

# Estimating Nitrogen Sufficiency Index Using A Natural Local Reference Approach

Yacine Bouroubi  
Effigis Geo-Solutions  
Montreal, Qc., Canada

Nicolas Tremblay, Philippe  
Vigneault, Carl Bélec  
Horticultural R&D Centre  
Agriculture and Agri-Food Canada  
St-Jean-sur-Richelieu, Qc., Canada

Viacheslav Adamchuk  
Department of Bioresource  
Engineering  
McGill University  
Ste-Anne-de-Bellevue, Qc., Canada

**Abstract**— Information-based crop management, such as variable rate technology, allows for changing rates of fertilization according to local needs. Fertilizer prescription maps can be derived from crop growth status assessed through proximal canopy sensing technologies. Typically, a vegetation index, such as the NDVI, is used to estimate a *Nitrogen Sufficiency Index* ( $NSI = NDVI / NDVI_{N-rich}$ ). NSI is related to N intake in comparison to the measurements in N-rich reference areas created artificially with high N rates applied at planting. This approach requires N-rich areas representative of different field conditions, which is not always practical. To circumvent this limitation ‘naturally N-rich’ reference areas linked to natural field characteristics can be identified. The goal of this present work was to show that the  $NSI_{NLR}$  estimated using less laborious Natural Local References (NLR) has a benefit comparable to the more traditional  $NSI_{N-rich}$  from N-rich plots. Experiments were performed on corn fields located in the Montérégie area of Quebec. The NDVI values were measured using 5 tractor-mounted GreenSeeker sensors. The spatial variability of the soil was analyzed using maps of apparent electrical conductivity ( $EC_a$ ) obtained using a Veris 3100 system. The  $NSI_{NLR}$ , set at the 90<sup>th</sup> percentile of the NDVI values obtained around each sensed area, was compared to  $NSI_{N-rich}$ . The NSI values estimated using both techniques were strongly correlated and demonstrated a high capacity to estimate the optimal N rate for corn.

**Keywords**—nitrogen sufficiency index; virtual reference, optimal nitrogen rate; proximal active sensors.

## I. INTRODUCTION

Nitrogen fertilizer (N) is an essential soil nutrient for food production. However, crop N needs and soil N availability are spatially and seasonally variable and the optimal N fertilization rate ( $N_{opt}$ ) differs within and between fields. When fixed application N rates are used, either crop stress (insufficient N supply) or environmental pollution (excess N supply) may result. Determining  $N_{opt}$  must take into account the spatial variability of the available soil N (by soil measurements) and crop N needs (by plant status observation). Current technology offers multiple solutions to measure efficiently (in terms of time and cost) soil properties and crop N status and to apply spatially-variable in-season N rates according to local needs. This leads to a control of over- and

under-applications of N and to improvements of N use efficiency (NUE).

Apparent soil electrical conductivity ( $EC_a$ ) is recognized as a useful technology for indirect measurement of important soil characteristics such as texture, water content, salinity, and general fertility ([1], [2]). Proximal active optical sensors, such as the GreenSeeker (Trimble LLC, Sunnyvale, California, USA), CropCircle (Holland Scientific, Inc., Lincoln, NE, USA), OptRx (AgLeader Technology, Inc., Ames, Iowa, USA), or CropSpec (Topcon Corporation Tokyo, Japan), measure canopy greenness and biomass and can be used to assess crop N requirements which is a major determinant of crop productivity ([3], [4]). Operationally, these sensors can be mounted on N fertilizer applicators equipped with computer processing and variable rate controllers to permit sensing and fertilization are done in one pass [5]. However, measurements from these sensors are influenced by undesired effects such as sensor properties, soil type, seasonal conditions, growth stage and cultivar specificity and could not be used directly for N status assessment ([6], [7]). Thus, the use of a reference N-rich plot (or strip) established with a high level of N applied since sowing is a useful benchmark for the sensed crop. Measurements from the two areas are used to calculate a nitrogen sufficiency index (NSI: NDVI of a sensed crop / NDVI of N-rich plot) which is then used to represent the potential need for additional N ([8], [9]). When the NSI approaches 1, the sensed crop looks as good as the N-rich crop and little additional N is probably required for the rest of the season. Conversely, low NSI indicates N-stress that could be corrected by the application of a higher rate of additional fertilizer. According to this principle, in-season N recommendation models based on NSI have been developed. The general premise for these models is that the greater the difference in sensor measurements between the sensed crop and the N-rich area (i.e. the lower the NSI), the more N fertilizer is needed ([10], [11], [5]). A review of NSI definitions and research history is given by [12].

The problem with this approach is that it requires N-rich areas representative of different field conditions, which is not realistic for commercial fields. Shanahan [5] used small

reference plots representing different soil conditions. A virtual reference approach that does not require any N-rich strip was originally proposed by [12]. In this approach, plants that are deemed to be as non-N limiting as those found in an artificially created N-rich strip are identified and their vegetation index values are used as the reference when calculating the NSI. This reference was fixed to the 95<sup>th</sup> percentile of CI (Chlorophyll Index) values obtained within a replication, with no relationship to soil properties.

The objective of this research was to 1) evaluate a strategy to define NSI based on natural N-rich field areas around each location of interest, 2) apply supplementary EC<sub>a</sub> data to assure relative soil homogeneity within fixed-radius areas around each location, and 3) evaluate the effect of the NSI estimation error on recommended optimum nitrogen application rate. Once successful, this approach could be used to enhance existing natural (or virtual) reference methods using supplementary soil EC<sub>a</sub> data.

## II. MATERIALS AND METHODS

### A. Fields properties and experimental design

Experiments were conducted in 2005, 2006, 2007, 2008, 2011 and 2012 in commercial corn (*Zea mays* L.) fields located in the Montérégie region of Quebec, Canada. Two fields were studied in 2012 (2012D and 2012S). Descriptions of the fields are given in Table 1. The soil were characterised by various textures of clay, clay-loam, loam, sandy loam and loamy sand. Soil variability was estimated by apparent electrical conductivity surveys (EC<sub>a</sub>, 0-30 cm) measured on bare soils using a Veris 3100 sensor cart system. Low EC<sub>a</sub> values corresponded to coarser soil textures and high EC<sub>a</sub> values, to finer soil textures. EC<sub>a</sub> was used in this instance to verify soil properties of the reference when calculating NSI values. In 2005 to 2008, N-rich strips were established with 250 kg N ha<sup>-1</sup> broadcast at sowing. In 2011 and 2012, an N-rich rate was applied only in limited areas selected according to three EC<sub>a</sub> classes (low, medium and high). This strategy was used to give an N-rich reference representing all soil conditions (Figure 1). The remaining areas received 30 kg N ha<sup>-1</sup> at sowing followed by one of 5-6 different in-season N rates (from 0 to 160 kg N ha<sup>-1</sup>). Yield responses to these N rates were used to estimate an optimal nitrogen fertilization rate for two NSI classes (high and low, separated at the median) and to demonstrate the value of NSI in assessing N fertilizer needs.

TABLE 1: Experimental fields descriptions.

Year	Location (UTM 18T)	Field size (ha)	Sowing date	Date of GreenSeeker observation and V stage
2005	629589 E, 4994395 N	8.7	12 May	28 June, V7
2006	629781 E, 4994184 N	6.5	9 May	5 July, V6-7
2007	629598 E, 4994275 N	8.7	3 May	22 June, V6
2008	629788 E, 4994156 N	13	6 May	25 June, V5-6
2011	649237 E; 4999019 N	9.4	6 June	29 June, V5

2012D	651970 E; 5001590 N	10.24	6 May	16 June, V5
2012S	649208 E; 4994185 N	11.85	19 May	18 June, V5

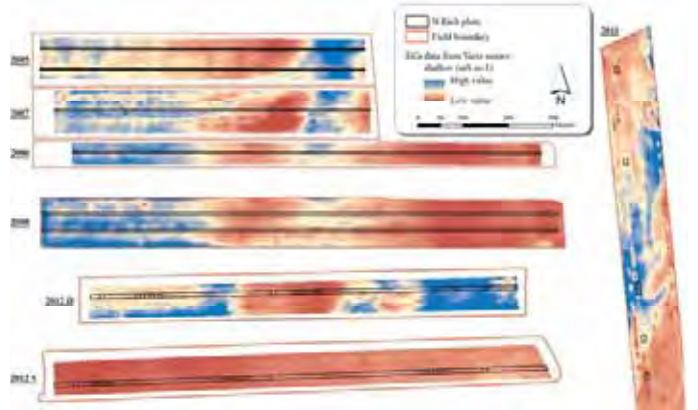


Fig. 1. Interpolated (ordinary kriging) EC<sub>a</sub> maps for the experimental fields.

### B. NDVI measurement and NSI calculation

NDVI measurements (Figure 2) were obtained before in-season N application from 5 GreenSeeker sensors (formerly NTech Industries, Ukiah, California, USA) mounted on a trailer pulled by a tractor. The NSI for every point was calculated using both N-rich strip and NLR. Since averaging NDVI<sub>N-rich</sub> for the entire strip appeared to be inappropriate due to the spatial variation of crop status measurements within a strip, the 90<sup>th</sup> percentile of the 20 nearest neighbours in the N-rich strip were used to define NDVI<sub>N-rich</sub> in 2005 through 2008. In 2011 and 2012, NDVI<sub>N-rich</sub> was defined as the 90<sup>th</sup> percentile of N-rich small plots grouped according to EC<sub>a</sub>. The reference NDVI<sub>NLR</sub> was set at the 90<sup>th</sup> percentile of NDVI values found in a local neighbourhood defined by a radius that should assure presence of naturally N-rich areas and at the same time provide comparable soil conditions. To assure these local soil homogeneity condition, NDVI values within the radius of search that represented different soil conditions (the difference dEC<sub>a</sub> < 2) were excluded from the NDVI<sub>NLR</sub> calculation.

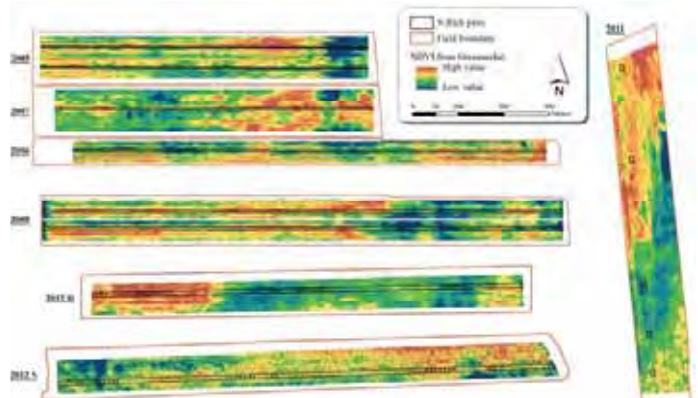


Fig. 2. Interpolated (ordinary kriging) NDVI maps from GreenSeeker sensors.

### C. Statistical analysis

Performance of the Natural Local References approach was assessed according to three criteria: 1) the correlation between  $NSI_{NLR}$  and  $NSI_{N-rich}$  (with a radius of NLR varying from 7 to 1100 m); 2) the similarity in soil type between the sensed area and the neighbourhood reference (small  $dEC_a$ ); and 3) the consequence of the quality of  $NSI_{NLR}$  estimates for  $Nopt$  prediction. The latter criterion was based on the calculation of  $dNopt = Nopt[low\ NSI] - Nopt[high\ NSI]$ , where  $Nopt$  was determined from a quadratic and plateau fitting of the yield response to N fertilization trials (Solari et al., 2010). The difference  $dNopt$  was calculated for  $NSI_{NLR}$  (with variable radius) and compared with the one obtained for  $NSI_{N-rich}$ . Comparisons were also made according to soil properties ( $Nopt[high\ EC_a] - Nopt[low\ EC_a]$ ). The aim of this analysis was to determine the best radius to be used in the NLR approach and to assess the efficiency of the proposed method.

## III. RESULTS AND DISCUSSION

### A. Comparison of $NSI_{VRP}$ with NDVI and $NSI_{N-rich}$

Figure 3a (specifically sub-figure a-1) shows the effect of soil  $EC_a$  on the  $NSI_{N-rich}$  vs NDVI relationship. Lower and higher NDVI values given by high and low  $EC_a$  classes, respectively, are raised to the same level when calculating  $NSI_{N-rich}$ . However, a strong correlation remains between NDVI and  $NSI_{N-rich}$  for the fields with wide N-rich strips (2005 to 2008) and for the field with low soil variability (2012S). In the fields 2011 and 2012D,  $NSI_{N-rich}$  calculated from small N-rich areas (that were supposed to represent soil types identified by the three  $EC_a$  classes) were not well correlated with NDVI; possibility, this was the result of unreliable  $NDVI_{N-rich}$  values. Figure 3a also shows that  $NSI_{NLR}$  tends to be highly correlated with NDVI for large radius since  $NDVI_{NLR}$  tends to be a unique value (90<sup>th</sup> percentile of the NDVI of the entire field) when the neighbourhood is too large.

The correlation between  $NSI_{NLR}$  and  $NSI_{N-rich}$  (Figure 3b) is high for the fields with N-rich strips (2005 to 2008) and for the field with low soil variability (2012S). This correlation is low for the fields 2011 and 2012D where  $NSI_{N-rich}$  were calculated from small N-rich areas. The radius of NLR approach gave the highest correlations between  $NSI_{NLR}$  and  $NSI_{N-rich}$  varying from 20 to 60 m for the fields studied in 2005 to 2008. This spatial extent is large enough to find a naturally N-rich area and close enough to ensure that the sensed area and its neighbourhood reference are in a comparable soil type. When soil is more homogeneous (2012S), this correlation is still high when the reference is taken from long separation distances (radius = 200 to 300 m).

### B. Natural Local References locations and soil properties

The spatial distribution of the Natural Local References (NLR) are shown (example of 2005) in figure 4 for radius extents of 30 and 60 m. When the N-rich strip was maintained in the search radius, much of NLR were identified inside this strip. When the N-rich strip was excluded (with a 4 m buffer to avoid border effects), NLR found for all GreenSeeker measurements (50,000 to 100,000 points, depending on the

field) were localized areas more limited in number when the radius was higher (60 m).

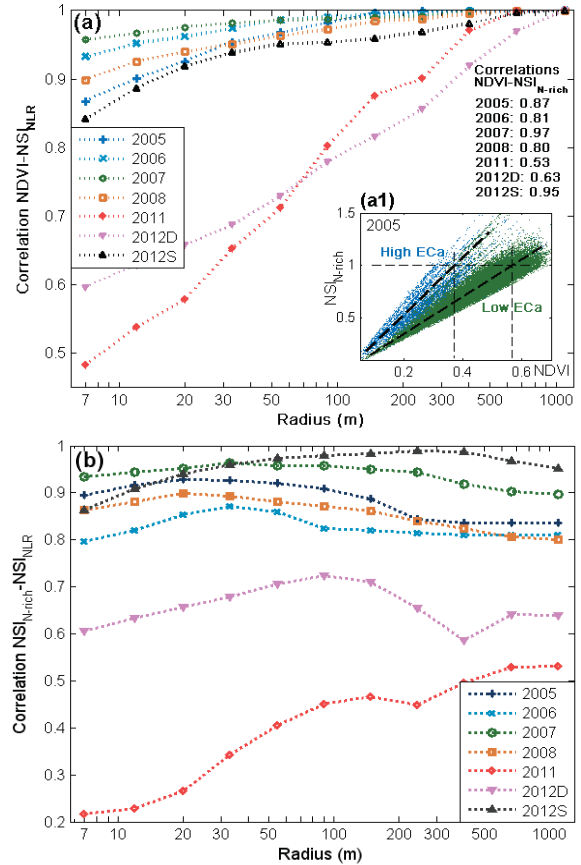


Fig. 3. Correlations NDVI- $NSI_{N-rich}$  (a1), NDVI- $NSI_{NLR}$  (a) and  $NSI_{N-rich}$ - $NSI_{NLR}$  (b) for different radius values (in log axis).

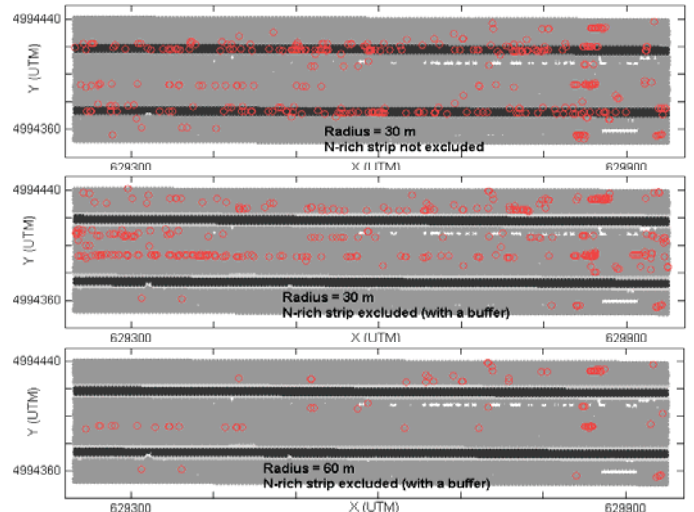


Fig. 4. Example (2005 field) of Neighbourhood References spatial distribution for two radius values.

The similarity between soil properties of the sensed area and the NLR used to calculate  $NSI_{NLR}$  were assessed through the difference of their  $EC_a$  ( $dEC_a$ ). When a radius  $<100$  m was used, most of  $dEC_a$  were within the  $\pm 2\ mS\ m^{-1}$  tolerance



interval to consider similar soil type and few points in the radius are excluded from the  $NDVI_{NLR}$  calculation (figure 5).

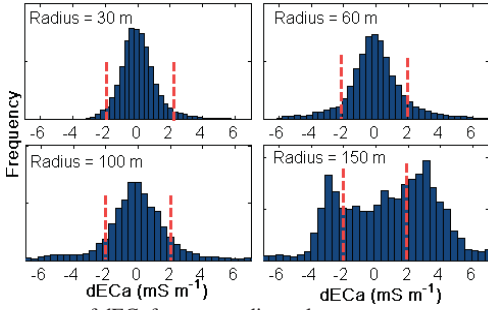


Fig. 5. Histograms of  $dEC_a$  for some radius values.

### C. Utility of $NSI_{NLR}$ for $N_{opt}$ prediction

As reported by [10], NSI can be used (in combination with  $EC_a$  when soil is variable) to help determine the optimal nitrogen fertilization rate ( $N_{opt}$ ). This section of the paper demonstrates the usefulness of  $NSI_{NLR}$  (as well as  $EC_a$ ) to better estimate  $N_{opt}$  through a quadratic fitting of Yield–N rate relationship. Figure 6 shows a typical example of this relationship (2005 field) where classes (low and high) of soil  $EC_a$ ,  $NSI_{NLR}$  and their combinations lead to considerable differences in  $N_{opt}$  ( $dN_{opt}$ ). In situation of low  $EC_a$  (coarse soil textures) or high  $NSI_{NLR}$  (N sufficiency), less N fertilizer is needed than in the opposite situation (high  $EC_a$  i.e. fine soil textures and low  $NSI_{NLR}$  i.e. N stress). The effects of combining both  $EC_a$  and  $NSI_{NLR}$  classes on  $N_{opt}$  will be analysed in a future research.

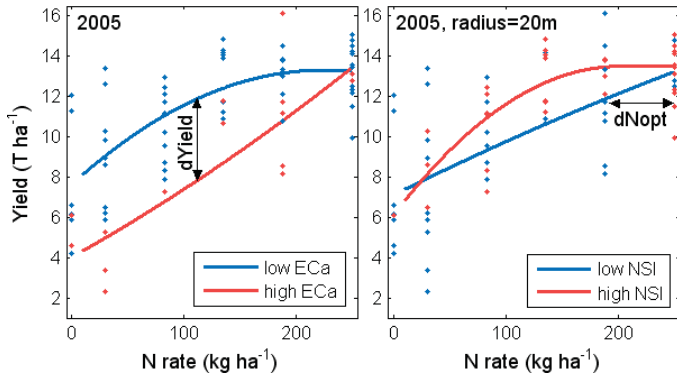


Fig. 6. Yield response to N rate for classes (high and low) of  $EC_a$ ,  $NSI_{NLR}$  (with a radius of 20 m) and their combinations for 2005 field.

Figure 7 shows that the best potential to estimate  $N_{opt}$  (highest  $dN_{opt}$ ) is obtained for an  $NSI_{NLR}$  corresponding to radius extending from 20 to 60 m according to years, except for 2012S field characterized by a homogeneous soil. Also, the use of  $NSI_{NLR}$  leads to  $dN_{opt}$  values comparable to those obtained with  $NSI_{N-rich}$  when a wide N-rich strip was used (2005 to 2008) and better than those obtained when N-rich references were limited to small areas (Table 2). Values of  $dN_{opt}$  vary from 30 to 100  $kg\ ha^{-1}$ , according to years. Note also that raw NDVI ( $NSI_{NLR}$  for very high radius) was also

useful for  $N_{opt}$  identification in 2005 to 2008 but not in 2011 and 2012D (Figure 7). Thus, results obtained for 2011 and 2012D fields are consistent with [5] remarks noting that without a reference, there is little basis for making N rate recommendations. The comparison of  $dN_{opt}$  obtained from the use of  $EC_a$  and NSI (Table 2) shows that NSI often leads to higher  $dN_{opt}$  than  $EC_a$ , but the latter gives generally higher contrasts ( $dYield$ ) in N response (as shown in figure 6 for the 2005 field). More investigations about this later result are necessary to clarify the situation in combining soil heterogeneity information and NSI approach for practical use.

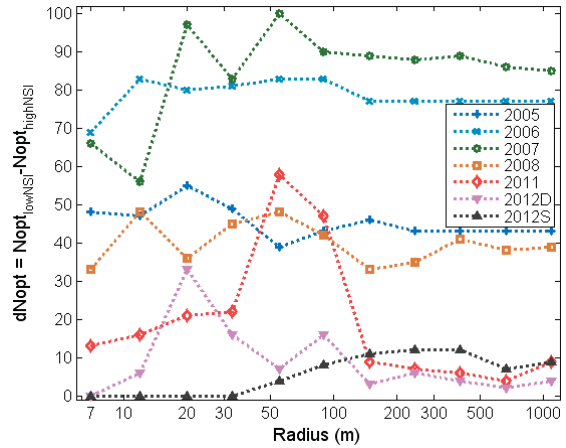


Fig. 7. Difference between  $N_{opt}$  ( $dN_{opt}$ ) of low and high  $NSI_{NLR}$  calculated with varying radius extents.

TABLE 2. Comparison of  $dN_{opt}$  obtained with high and low classes of  $EC_a$ ,  $NSI_{N-rich}$  and  $NSI_{NLR}$  (with the best radius) for all used fields.

$dN_{opt}$ ( $kg\ ha^{-1}$ )	2005	2006	2007	2008	2011	2012D	2012S
from $EC_a$	32	60	133	24	40	15	-
from $NSI_{N-rich}$	46	89	116	13	-6	0	9
from $NSI_{NLR}$	55	84	100	48	58	33	12
Best radius (m)	20	55	55	55	55	20	400

## IV. CONCLUSION

The Natural Local References (NLR) approach proposed in this paper offers a way to calculate the optimal N rate estimation. Nitrogen sufficiency index calculated using this approach could be derived from remote sensing images or from active proximal canopy sensors. In high NSI conditions, important N savings (up to 100  $kg\ ha^{-1}$ ) in corn fertilization could be made resulting in high economic and environmental benefits. Same findings were reported by [4] who showed that the savings resulting from the use of NSI reach levels as high as 100 lb/A. The latter authors as well as [11] recommend reducing N fertilizer rate also for very low NSI values (back-off principle).

The NLR approach is suitable for use in N management strategies based on NSI like those developed by [5] and [10]. When proximal sensing is combined with fertilizer application



equipment is used, the practical implementation of the NLR approach can be based on 'Drive-First' or 'Drive-and-Apply' principles, as terms used by [12], who showed that the two approaches give close N rates after a certain distance covered by the sensor. In the second case ('Drive-and-Apply'), since the NLR based on a particular radius could not be established all over the field, one can take as a reference the 90<sup>th</sup> percentile of the measurements taken 20 to 50 meters before the sensed area. As shown in this paper, a more complex algorithm involving spatial soil heterogeneity data could also be designed for this purpose.

#### REFERENCES

- [1] N. E. Derby, F. X. M. Casey, and D. W. Franzen, "Comparison of nitrogen management zone delineation methods for corn grain yield," *Agronomy Journal*, vol. 99(2), pp. 405–414, 2007.
- [2] N. Tremblay, Y. Bouroubi, P. Vigneault, and C. Belec, "Guidelines for in-season nitrogen application for maize (*Zea mays* L.) based on soil and terrain properties," *Field Crops Research*, vol. 122, pp. 273–283, 2011.
- [3] N. Tremblay, Z. J. Wang, B. L. Ma, C. Belec, and P. Vigneault, "A comparison of crop data measured by two commercial sensors for variable-rate nitrogen application," *Precision Agriculture*, vol. 10, pp. 145–161, 2009.
- [4] D. F. Roberts, N. R. Kitchen, P. C. Scharf, and K. A. Sudduth, "Will variable-rate nitrogen fertilization using crop canopy reflectance sensing deliver environmental benefits?," *Agronomy Journal*, vol. 102(1), pp. 85–95, 2010.
- [5] J. Shanahan, "Using Crop Sensors to Improve Corn Nitrogen Management," *Crop Insights*, vol. 20(6), 2010.
- [6] N. Tremblay, "Determining nitrogen requirements from crops characteristics. Benefits and challenges," *Recent Res. Devel. Agronomy & Horticulture*, vol. 1, pp. 157–182, 2004.
- [7] J. J. Schroder, J. J. Neeteson, O. Oenema, and P. C. Struik, "Does the crop or the soil indicate how to save nitrogen in maize production?: Reviewing the state of the art," *Field Crops Research*, vol. 66(2), pp. 151–164, 2000.
- [8] T. M. Blackmer, and J. S. Schepers, "Techniques for monitoring crop nitrogen status in corn," *Communications in Soil Science and Plant Analysis*, vol. 25(9), pp. 1791–1800, 1994.
- [9] N. Tremblay, and C. Bélec, "Adapting nitrogen fertilization to unpredictable seasonal conditions with the least impact on the environment," *Horttechnology*, vol. 16(3), pp. 408–412, 2006.
- [10] N. Tremblay, Y. Bouroubi, B. Panneton, S. Guillaume, and P. Vigneault, "Development and validation of a fuzzy logic estimation of optimum N rate for corn based on soil and crop features," *Precision Agriculture*, vol. 11, pp. 621–635, 2010.
- [11] K. H. Holland, and J. S. Schepers, "Derivation of a Variable Rate Nitrogen Application Model for In-Season Fertilization of Corn," *Agronomy Journal*, vol. 102 (5), pp. 1415–1424, 2010.
- [12] K. H. Holland, and J. S. Schepers, "Use of virtual-reference Concept to interpret active crop canopy sensor data," *Precision Agriculture*, vol. 14 (1), pp. 71–85, 2013.

# Field comparison of ultrasonic and canopy reflectance sensors used to estimate biomass and N-uptake in sugarcane

G. Portz<sup>1</sup>, L.R. Amaral<sup>1</sup>, J.P. Molin<sup>1</sup> and V.I. Adamchuk<sup>2</sup>

<sup>1</sup>Biosystems Engineering Department, University of São Paulo, 11 Pádua Dias Av., Piracicaba, SP, 13418-900, Brazil; [gportz@gmail.com](mailto:gportz@gmail.com)

<sup>2</sup>Department of Bioresource Engineering, McGill University, 21,111 Lakeshore Road, Ste-Anne-de-Bellevue, QC, H9X 3V9, Canada

## Abstract

The use of crop canopy reflectance sensors is being studied intensively in sugarcane crops; however, real-time measurements of plant height provide another possibility for estimating the spatial variability of biomass and nitrogen uptake by the crop during the in-season fertilization period. An ultrasonic proximity sensor system was deployed to measure crop canopy height. The data were combined with commercially available canopy optical reflectance measurements to characterize spatial variability of crop growth in two commercial areas (a sandy soil field and a clay soil field) during two different growing stages. After mapping, ten points in each field were defined to determine plant biomass and N-uptake through manual sampling and traditional measurements. Through the data analysis, a pixel by pixel comparison was performed to relate the interpolated maps obtained using different sensor systems. The ten points in each field were used to relate the actual biomass and N-uptake with the sensor data. It was shown that all correlations were significant and there was a slight indication that canopy reflectance sensing produced a better assessment of crop growth at the earlier growth stage whereas the ultrasonic sensor resulted in more accurate predictions at the later growing stage.

**Keywords:** ultrasonic sensor, canopy height, plant reflectance, biomass, N-uptake

## Introduction

Sugarcane (*Saccharum spp.*) is the most important crop for the production of sugar and ethanol in tropical and subtropical regions. As with most grass species, nitrogen (N) is one of the major inputs for this crop; nevertheless, there is no reliable and inexpensive soil analyses procedure to determine N availability *in situ* during the growing season on tropical soils. To manage nitrogen fertilization according to local needs, canopy reflectance sensors are already being intensively studied and results show that they can provide the status of biomass and N nutrition of sugarcane in real time (Amaral *et al.*, 2012, Portz *et al.*, 2012a). However, these sensors exhibit signal saturation when the sugarcane crop is above the 0.6 m stem height. As shown previously (Portz *et al.*, 2012b), increasing biomass does not reflect an adequate increase of the vegetation index (VI) later in the season when there is still room to conduct nitrogen fertilization in-season.

As an alternative, measuring crop canopy using ultrasonic proximity sensors is not new. During the 1980s, Shibayama *et al.* (1985) used such a sensor for crop canopy characterization and Sui *et al.* (1989) developed a measurement system using ultrasonic ranging modules to provide cumulative plant volume, plant width and height for a variety of bush-type plants, such as cotton and soybean. Later, ultrasonic proximity sensors were used simultaneously with crop canopy reflectance sensors. Scotford and Miller (2003) concluded that ultrasonic sensors proved useful for monitoring winter wheat growth beyond GS 30, whereas NDVI measurements were useful for monitoring winter wheat up to GS 31 (before the point of stem elongation). This evidence suggests that by combining these two measurements, the crop can be monitored throughout the entire growing season.



Sui and Thomasson (2006) combined plant reflectance sensors and ultrasound sensors to determine the status of nitrogen in cotton. The results showed that the spectral information and plant height measured by the system showed significant correlation with the nitrogen contained in the leaves of the cotton plants.

Shrestha *et al.* (2002) investigated ultrasonic sensor estimates of corn plant height in a lab environment. They showed that the estimated height correlated with the manually measured height. While working with an optical reflectance sensor on corn, Freeman *et al.* (2007) also took measurements of plant height during the investigation and concluded that plant height alone was a good predictor of plant biomass at all stages of growth sampled. Similarly, recent work by Shiratsuchi *et al.* (2009) showed that correlations between height, biomass and nitrogen were observed in corn, suggesting the integration of plant reflectance sensor and ultrasound measurements of plant height, to estimate nitrogen extracted by the crop during the entire growing season.

During previous work, Portz *et al.* (2012c) observed that sugarcane stem height measured manually correlated with the crop biomass and N-uptake. This means that crop canopy height can provide complementary information or it can be considered an alternative to optical reflectance sensing when estimating the biomass and N-uptake in the crop. The objective of this project was to compare canopy reflectance and ultrasonic crop height sensing when predicting sugarcane biomass and N-uptake in real time.

### Materials and methods

The study involved two commercial fields located in the state of São Paulo, Brazil, on sandy (12 ha) and clay (16 ha) soils. The fields were harvested during mid-season (first ratoon), and were mapped twice, when the crop was at 0.3 m to 0.5 m (11/10/2012), and at 0.5 m to 0.7 m (14/11/2012) average stem height. The fields were scanned simultaneously with a commercial crop canopy reflectance sensor (N-Sensor™ ALS, Yara International ASA, Research Centre Hanninghof, Duermen, Germany), and two ultrasonic sensors (both Polaroid 6500, Minnetonka, Minnesota, USA). The N-Sensor ALS is comprised of a transmitter with a xenon flashlight, providing high intensity illumination between 650 and 1,100 nm and a 10 Hz receiver with 2 photodiodes and interference filters of 730 and 760 nm in front of them measuring the proprietary defined vegetation index (Jasper *et al.*, 2009). Two ultrasonic sensors were operated at a frequency of 49.4 kHz and connected to a data logger (CR 1000 Campbell Scientific, Logan, Utah, USA). All of the measurements were geo-referenced using a Global Positioning System (GPS) receiver and a 1 Hz log file was created by synchronizing and aggregating all of the data. The sensors were mounted on top of a high clearance vehicle (Figure 1) that passed the field every 10 rows (15 m) at a travel speed between 12 and 15 km/h. This produced about 200 independent data points per hectare. Since the crop canopy reflectance sensor had an



Figure 1. Canopy sensors disposition on a high clearance vehicle.



oblique view, six crop rows were sensed at a time. Each ultrasonic system evaluated only one row, while being installed vertically.

The average of the recorded data from the left and right sensors were filtered, cutting off points outside the field boundaries and the negative values; they were interpolated (inverse distance weighting) using a 5×5 m raster. A five-color legend was applied for visual analysis. Ten validation locations were selected in each field to represent the entire range of sensor-based measurements for both systems. Each validation location was in the middle of a 5 m cell. In every case, destructive plant samples of the above ground biomass were taken by manually cutting a 1.5 m sub-plot consisting of three rows (4.5 m). The samples were weighed in the field, and processed in the laboratory to measure total N content (Kjeldahl method) and, ultimately, the crop N-uptake. Figure 2 illustrates the entire process. The sensor-based data were compared using a regression analysis applied to all of the interpolated pixels. Another set of linear regression analyses was conducted to relate the measured biomass, the N-uptake and the sensor measurements.

### Results and discussion

Figure 3 presents the interpolated data for the canopy reflectance and the canopy height for the two measurements made on the sandy soil field. All of the maps illustrated the same spatial pattern with slightly less small-scale variability in crop canopy reflectance maps due to the smoothing caused by the oblique position of the sensor. The same conclusions can be made while observing the maps produced based on the data collected from the clay soil field (Figure 4).



Figure 2. Sampling process.

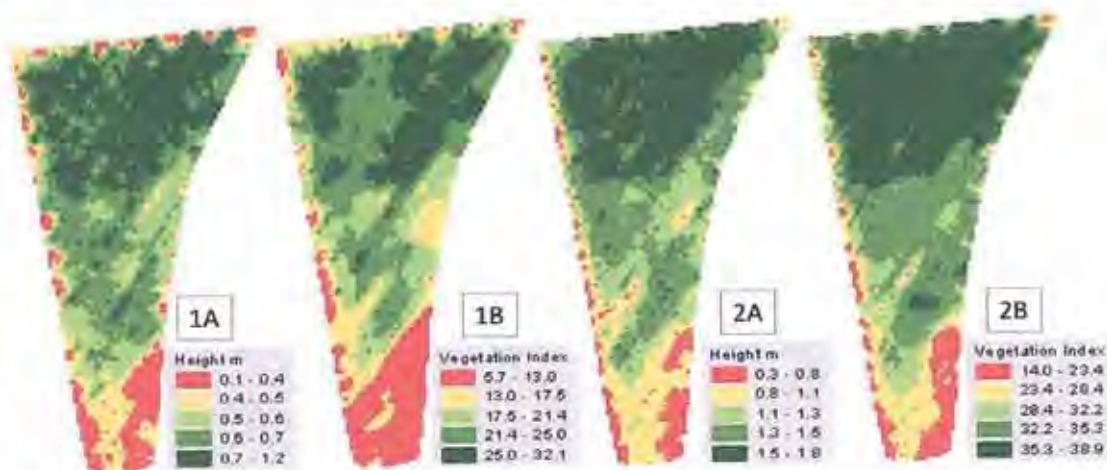


Figure 3. Interpolated maps of crop height during the first (1A) and second (2A) mapping and crop canopy reflectance during the first (1B) and second (2B) scanning of the sandy soil field.

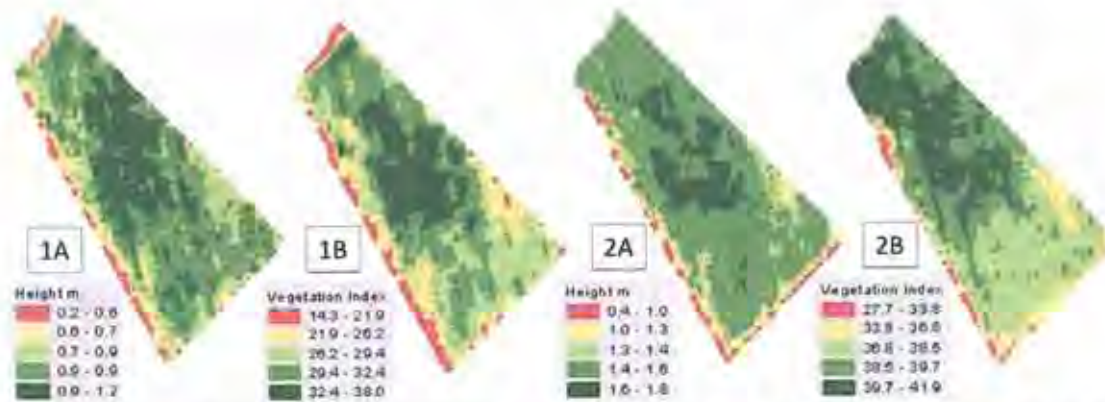


Figure 4. Interpolated maps of crop height during the first (1A) and second (2A) mapping and crop canopy reflectance during the first (1B) and second (2B) scanning of the clay soil field.

A pixel-by-pixel comparison of corresponding maps produced using two different sensor systems (Figure 5) showed a strong, but field and growth stage specific, correlation. This is consistent with previously published observations (Portz *et al.*, 2012a). Naturally, it is possible to see that during the first scanning of the sandy soil field (04 VI - 32 VI), the crop was shorter (0.2-0.9 m) when compared to taller crops (0.4-1.8 m) during the second scanning (16-38 VI). The crop canopy reflectance measurements did not exhibit this change due to possible sensor saturation when the canopy became closed. Furthermore, it could be seen that the range of measurements conducted by the ultrasonic sensor increased during the second scan indicating a strong sensor response to crop biomass. On the other hand, especially in the case of clay soil, the range of vegetation index values has declined at the later growth stage indicating that the sensor was less capable of distinguishing crop performance in different parts of the field when compared to the earlier growing stage. This could be explained by the fact that optical measurements primarily observe crop leaves and height

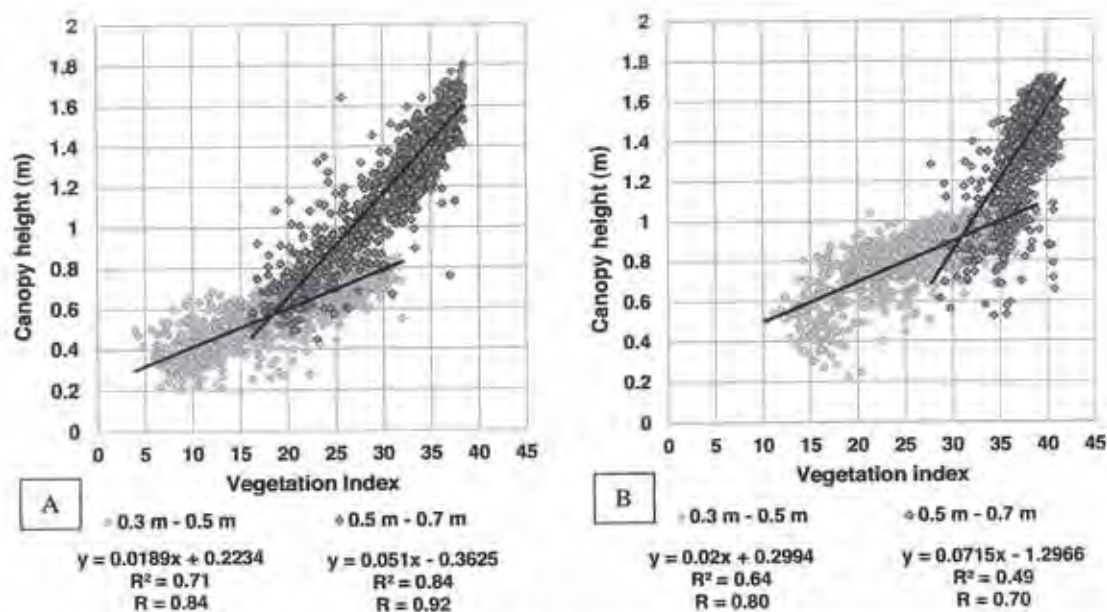


Figure 5. Pixel by pixel correlation of first and second scanning moments on sandy (A) and clay (B) soil fields.



estimates assess the stalks. Thus, during the first scanning over a younger sugarcane crop, the leaves were more relevant to crop biomass and N uptake than the stalks. During the second scanning, the stalk was a bigger contributor to crop biomass leaving crop height measurements more responsive to spatially different crop performance.

Through the analysis of data obtained from validation locations, it follows that sensor-based measurements correlated to both biomass and N-uptake in every field and during each scan (Figure 6). Based on the analysis of regression models (Table 1), it once again appears that both sensor systems responded to changes in crop biomass and N uptake with coefficients of determination ( $R^2$ ) at, or above, 0.70. Root mean squared error (RMSE) for biomass ranged between 1.66 and 6.25 Mg/ha and N-uptake varied between 4.8 and 10.4 kg/ha. Except for the height-based N-uptake estimates in the field with clay soil, RMSE values were higher during the second mapping event. In terms of the two sensor systems, in every case, the estimated RMSE were lower for crop canopy reflectance sensors during the first scanning and for the ultrasonic sensors during the second scanning. Many of these differences were not significant statistically and further research is needed to assess the quality of the predictions in different growing conditions and to verify if combining both sensors would reveal any benefits when compared to using one of them at a time.

## Conclusions

The sensor data presented consistent correlations when generating similar maps for canopy reflectance and canopy height on the two studied areas. Sampled biomass and N-uptake also correlated with both canopy reflectance and crop height sensor-based measurements. With very little statistical significance, crop canopy reflectance measurements seemed more responsive to crop biomass and N uptake at a younger growth stage. Ultrasonic sensor crop height measurements seemed to be more relevant to more developed crops. This result indicates that it might be beneficial to integrate both sensors to make measurements during the entire growing season. More research is needed to evaluate the statistical significance of these findings and explore the benefits of such an integrated system applied to a sugarcane growing environment.

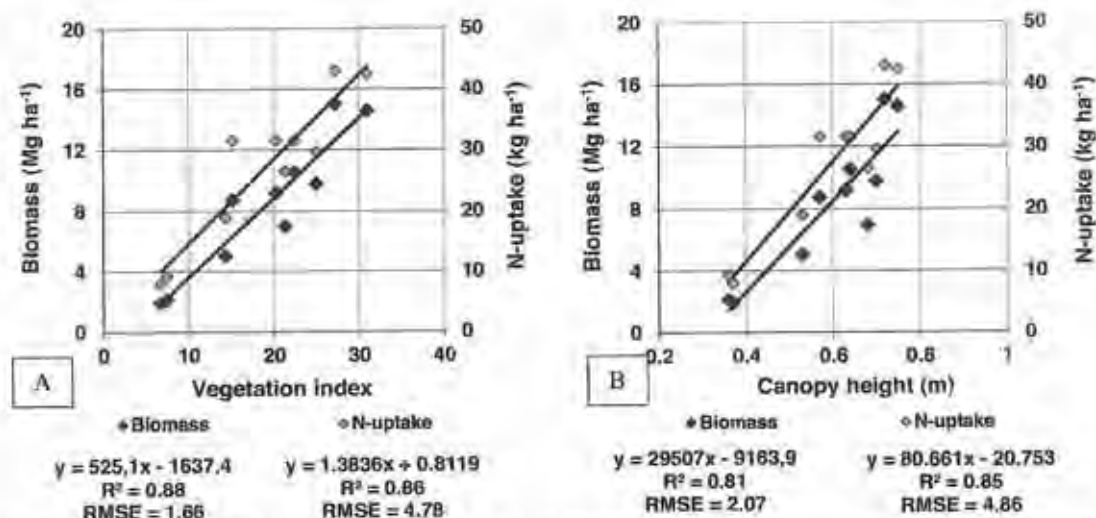


Figure 6. Canopy reflectance (A) and canopy height (B) related to biomass and N-uptake of the sample points from the first scanning over the sandy soil field.



Table 1. Analysis of relationships biomass and N-uptake as functions of vegetation index and crop height measurements.

Sensor	Value	Soil	Mapping	Intercept <sup>1</sup>	Slope <sup>1</sup>	RMSE <sup>2</sup>	R <sup>2</sup>
<b>Biomass (Mg/ha)</b>							
Canopy reflectance	Vegetation index	Sand	First	-1.6 <sup>NS</sup>	0.53 <sup>***</sup>	1.66 <sup>abc</sup>	0.88
			Second	-29.9 <sup>*</sup>	1.63 <sup>**</sup>	6.25 <sup>efg</sup>	0.75
		Clay	First	-14.2 <sup>*</sup>	0.97 <sup>**</sup>	2.72 <sup>abcde</sup>	0.80
			Second	-211.6 <sup>**</sup>	6.49 <sup>**</sup>	5.84 <sup>defg</sup>	0.81
Ultrasonic sensing	Height (m)	Sand	First	-9.2 <sup>*</sup>	29.51 <sup>***</sup>	2.07 <sup>abcd</sup>	0.81
			Second	-18.5 <sup>*</sup>	32.39 <sup>***</sup>	5.27 <sup>defg</sup>	0.82
		Clay	First	-38.8 <sup>*</sup>	58.03 <sup>**</sup>	3.42 <sup>bcdef</sup>	0.69
			Second	-65.4 <sup>**</sup>	71.21 <sup>***</sup>	4.69 <sup>cdefg</sup>	0.88
<b>N-uptake (kg/ha)</b>							
Canopy reflectance	Vegetation index	Sand	First	0.8 <sup>NS</sup>	1.4 <sup>***</sup>	4.78 <sup>abc</sup>	0.86
			Second	-44.8 <sup>*</sup>	2.9 <sup>***</sup>	10.43 <sup>cde</sup>	0.78
		Clay	First	-24.8 <sup>NS</sup>	2.3 <sup>**</sup>	6.71 <sup>abcde</sup>	0.79
			Second	-259.6 <sup>*</sup>	8.7 <sup>**</sup>	9.61 <sup>cde</sup>	0.74
Ultrasonic sensing	Height (m)	Sand	First	-20.8 <sup>*</sup>	80.7 <sup>***</sup>	4.86 <sup>abcd</sup>	0.85
			Second	-24.3 <sup>*</sup>	57.9 <sup>***</sup>	8.64 <sup>bcde</sup>	0.85
		Clay	First	-85.9 <sup>*</sup>	141.6 <sup>**</sup>	8.05 <sup>abcde</sup>	0.70
			Second	-65.9 <sup>*</sup>	97.5 <sup>**</sup>	7.72 <sup>abcde</sup>	0.83

<sup>1</sup> Parameters of a simple linear regression significant at  $\alpha=0.05$  (\*),  $\alpha=0.01$  (\*\*),  $\alpha=0.001$  (\*\*\*), as well as not significant at  $\alpha=0.05$  (NS).

<sup>2</sup> RMSE estimates with common letters (a-g) are not significantly ( $\alpha=0.05$ ) different from each other.

## Aknowledgements

This work would not be possible without the collaboration of São Martinho's Mill team, Máquinas Agrícolas Jacto and Yara Hanninghof. We also acknowledge the Research and Projects Financing (FINEP) from the Ministry of Science and Technology through the PROSENSAP project for financial support and the National Council for Scientific and Technological Development (CNPq) for providing the doctoral scholarship to the first author.

## References

- Amaral, L.R., Portz, G., Rosa, H.J.A. and Molin, J.P. 2012. Use of active crop canopy reflectance sensor for nitrogen sugarcane fertilization. In: Proceedings of 10<sup>th</sup> International Conference on Precision Agriculture (ISPA), Indianapolis, USA (CD-ROM).
- Freeman, K.W., Arnall, D.B., Mullen, R.W., Girma, K., Martin, K.L., Teal, R.K. and Raun, W.R. 2007. By-plant prediction of corn forage biomass and nitrogen uptake at various stages using remote sensing and plant height measures. *Agronomy Journal* 99 (2) 530-536.
- Jasper, J., Reusch, S. and Link, A. 2009. Active sensing of the N status of wheat using optimized wavelength combination – impact of seed rate, variety and growth stage. In: Van Henten, E.J., Goense, D. and Lokhorst, C. (Eds.), *Precision agriculture '09. Proceedings of the 7<sup>th</sup> European Conference on Precision Agriculture*. Wageningen Academic Publishers, the Netherlands, pp. 23-30.

- Pörtz, G., Molin, J.P. and Jasper, J. 2012a. Active crop sensor to detect variability of nitrogen supply and biomass on sugarcane fields. *Precision Agriculture* 13 33-44.
- Portz, G., Amaral, L.R., Molin, J.P. and Jasper, J. 2012b. Optimum sugarcane growth stage for canopy reflectance sensor to predict biomass and nitrogen uptake. In: *Proceedings of 10<sup>th</sup> International Conference on Precision Agriculture, Indianapolis, USA (CD-ROM)*.
- Portz, G., Amaral, L.R., Molin, J.P. 2012c. Measuring sugarcane height in complement to biomass sensor for nitrogen management. In: *Proceedings of 10<sup>th</sup> International Conference on Precision Agriculture (ISPA), Indianapolis, USA (CD-ROM)*.
- Scottford, I.M. and Miller, P.C.H. 2003. Combination of spectral reflectance and ultrasonic sensing to monitor the growth of winter wheat. *Biosystems Engineering* 87 (1) 27-38.
- Shiratsuchi, L.S., Ferguson, R.B., Adamchuk, V.I., Shanahan, J.F. and Slater, G.P. 2009. Integration of ultrasonic and active canopy sensors to estimate the in-season nitrogen content for corn. In: *Proceedings of the 39<sup>th</sup> North Central Extension-Industry Soil Fertility Conference, 18-19 Norcross, Georgia, USA: International Plant Nutrition Institute*.
- Shibayama, M., Akiyama, T. and Munakata, K. 1985. A portable field ultrasonic sensor for crop canopy characterization. *Remote Sensing of Environment* 18 269-279.
- Shrestha, D. S., Steward, B.L., Birrell, S.J. and Kaspar, T.C. 2002. Corn plant height estimation using two sensing systems. Paper No. 021197, ASAE, St Joseph, MI, USA.
- Sui, R. Wilkerson, J.B., Wilhelm, L.R. and Tompkins, F.D. 1989. A microcomputer-based morphometer for bush-type plants. *Computer and Electronics in Agriculture* 4 (1) 43-58.
- Sui, R. and Thomasson, J.A. 2006. Ground-based sensing system for cotton nitrogen status determination. *Transactions of the ASABE* 49 1983-1991.

# Theoretical basis for sensor-based in-season nitrogen management

V.I. Adamchuk

McGill University, Ste-Anne-de-Bellevue, QC, H9X 3V9, Canada; [vjacheslav.adamchuk@mcgill.ca](mailto:vjacheslav.adamchuk@mcgill.ca)

## Abstract

Managing nitrogen fertilization of corn and other crops during the vegetation stages is a growing practice implemented to increase the efficient use of nitrogen. Active crop canopy sensing has been employed to adjust nitrogen application rates in response to the spatial variability of vegetation growth. Several different application algorithms have been developed to convert sensor measurements into optimized nitrogen application rates. While assuming a second-order polynomial and plateau crop response function, this paper illustrates the derivation of a decision-support function to account for changes in fertilizer and crop prices. Increases in the fertilizer-to-crop cost ratio tend to cause a negative offset to the recommended N application rate. With further evaluation in terms of uncertainties as well as the effects of soil and weather, these results can be used to develop profit-maximizing, sensor-based algorithms for in-season nitrogen management.

**Keywords:** crop canopy sensing, nitrogen fertilization, profitability

## Introduction

Studies of the relationship between leaf color, chlorophyll content, nitrogen supply, biomass, and grain yield in agronomic crops have taken place around the world. This intensive research produced a variety of crop canopy sensing instruments used to manage nitrogen (N) fertilization during the growing season. Each of these instruments involves photo-detectors and a light source (ambient light was used in some of the older systems). Currently, active crop canopy sensor systems use a combination of light reflectance measurements in the visible and near-infrared parts of the spectrum to calculate various vegetation indices that infer the growth status of a particular crop. A high correlation between canopy sensor information and the crop N status have been found for wheat, maize and rice. The most popular reason for using active canopy sensors is to optimize the use of nitrogen fertilizers during in-season applications (Kitchen *et al.*, 2010).

While commercial crop canopy sensor systems provide useful information pertaining to the status of the vegetation (vigor, color and amount of biomass), algorithms to convert sensor measurements into economically and environmentally optimized N application rates (Bullock and Bullock, 1994) are needed to take advantage of this information (Samborski *et al.*, 2009). The objective of this paper is to illustrate a mathematical derivation of a profit-maximizing nitrogen application rate as a function of relative sensor measurements, while assuming a second-order polynomial and plateau crop response to fertilization.

## Materials and methods

The first step in developing an algorithm for optical, sensor-based N fertilization is to define a combination of light reflectance measurements that corresponds to different spectral wavelengths providing the best indicators of crop performance at any given stage of growth (e.g. vegetation index). The normalized difference vegetation index (NDVI) and the chlorophyll index (CI) are just two examples of the most popular indices (Solari *et al.*, 2008; Shiratsuchi *et al.*, 2011).

Next, it is important to define a relative sensor indicator (RSI) that can be used as a proxy of the potential for N response to fertilization. For example, this could be a sufficiency index (SI) concept (Biggs *et al.*, 2002), when the real-time index value is normalized by the reference value obtained from



a sufficiently fertilized area under matching growing conditions. In general, RSI can be represented as a linear transformation that scales and offsets the vegetation index and can be defined as:

$$RSI = b_0 + b_1 \cdot VI \quad (1)$$

where:

$VI$  = vegetation index (NDVI, CI, etc.);

$b_0$  and  $b_1$  = linear transformation coefficients.

In the case of a sufficiency index concept ( $b_0=0$  and  $b_1=1/VI_{max}$ ) according to Holland and Schepers (2010):

$$SI = \frac{VI}{VI_{max}} \quad (2)$$

where:

$VI_{max}$  = vegetation index value from a sufficiently fertilized field area,

#### Yield-maximizing algorithm

Among different yield response models, a second-order polynomial and plateau model (Havlin *et al.*, 2004; Bongiovanni and Lowenberg-DeBoer, 2000; Solari *et al.*, 2008) represents an appealing form of response that is relatively easy to analyze and to interpret the results. In this case:

$$Y = \begin{cases} c_0 + c_1 N + c_2 N^2 & \text{for } N < N_{Ymax} \\ c_0 + c_1 N_{Ymax} + c_2 N_{Ymax}^2 & \text{for } N \geq N_{Ymax} \end{cases} \quad (3)$$

where:

$Y$  = crop yield in physical units;

$N$  = soil nitrogen available for crop in physical units;

$N_{Ymax}$  = the smallest value of  $N$  that corresponds to the maximum yield;

$c_0$ ,  $c_1$ , and  $c_2$  = coefficients of the second-order polynomial.

The maximum yield  $Y_{max}$  and corresponding  $N_{Ymax}$  is solved for when  $dY/dN = 0$ . As long as the relationship between anticipated crop yield and  $RSI$  is linear ( $Y = d_0 + d_1 \cdot RSI$ ), Equation 3 can be rewritten as an  $RSI$  response function (Figure 1):

$$RSI = \begin{cases} a_0 + a_1 N + a_2 N^2 & \text{for } N < N_{max} \\ a_0 + a_1 N_{max} + a_2 N_{max}^2 & \text{for } N \geq N_{max} \end{cases} \quad (4)$$

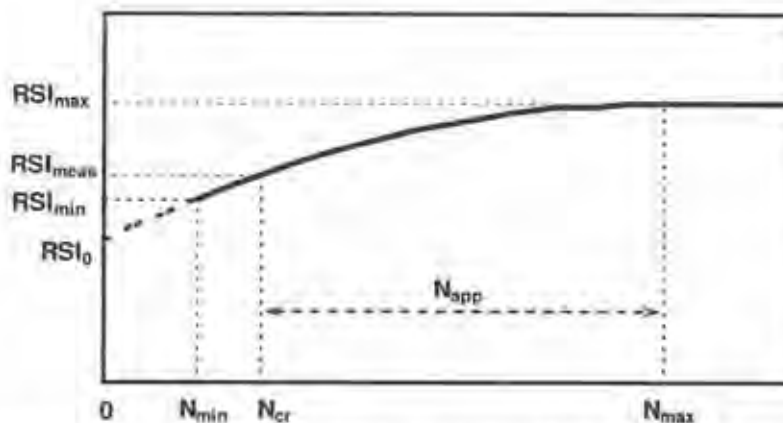


Figure 1. Second-order polynomial and plateau response of RSI to N availability.

where  $a_0$ ,  $a_1$ , and  $a_2$  = coefficients of the second-order polynomial that can be derived using  $c_{0p}$ ,  $c_1$ , and  $c_2$  as well as the parameters of linear relationship between yield and RSI. Similar to the yield, the maximum value of RSI ( $RSI_{max}$ ) is obtained when  $dRSI/dN = 0$ . Thus  $N_{max}$  can be defined using:

$$a_1 + 2a_2N_{max} = 0 \Rightarrow N_{max} = -\frac{a_1}{2a_2} \quad (5)$$

In the case of the sufficiency index concept,  $RSI_{max} = SI_{max} = 1$  when  $VI = VI_{max}$ . At the low end, zero N availability ( $n=0$ ) defines the  $RSI_0$ :

$$RSI_0 = a_0 \quad (6)$$

It is important to note that  $RSI_0$  ( $n=0$  in Figure 1) is not practical because crop growth is likely to be severely limited and the response function probably would not follow the same second-order polynomial model below a certain attainable  $RSI_{min}$  (check plot). Instead, it represents the point of intersection between the overall RSI response function and the axis of ordinate. Substituting values into Equation 4 using  $a_0$  from Equation 6 and  $a_1$  found by solving Equation 5 produces:

$$RSI_{max} = RSI_0 - 2a_2N_{max}^2 + a_2N_{max}^2 \Rightarrow a_2 = \frac{RSI_0 - RSI_{max}}{N_{max}^2} \quad (7)$$

At the time of the in-season N application, the measured RSI ( $RSI_{meas}$ ) value represents the crop response to an already available N credit ( $N_{cr}$ ). This availability can be attributed to various factors including soil conditions and the previous cropping system ( $N_{cr} > N_{min}$ ). Based on the quadratic response part of Equation 4:

$$RSI_{meas} = a_0 + a_1N_{cr} + a_2N_{cr}^2 \quad (8)$$

Solving this equation for the smaller of two values of  $N_{cr}$  produces:

$$N_{cr} = \frac{a_1 - \sqrt{a_1^2 - 4a_2(a_0 - RSI_{meas})}}{2a_2} = N_{max} - \sqrt{N_{max}^2 - \frac{RSI_0 - RSI_{meas}}{a_2}} \quad (9)$$

Using an assumption that at N availability is the sum of  $N_{cr}$  and N application ( $N_{app}$ ), the quadratic response part of Equation 4 can be rewritten as:

$$RSI_{res} = a_0 + a_1(N_{cr} + N_{app}) + a_2(N_{cr} + N_{app})^2 \quad (10)$$

where  $RSI_{res}$  = expected RSI value after N application.

Since  $N_{app}$  is the only manageable part of N, maximizing  $RSI_{res}$  means that  $dRSI_{res}/dN_{app} = 0$ . This produces:

$$\frac{dRSI_{res}}{dN_{app}} = a_1 + 2a_2N_{cr} + 2a_2N_{app} = 0 \quad (11)$$

Solving for  $N_{app}$ :

$$N_{app} = -\frac{a_1}{2a_2} - N_{cr} = N_{max} - N_{cr} \quad (12)$$

Substituting  $a_2$  in Equation 9 using Equation 7 and then  $N_{cr}$  in Equation 12 using Equation 9 results in the following equation, similar to previously described by Holland and Schepers (2010):

$$N_{app} = N_{max} - N_{cr} + \sqrt{N_{max}^2 + N_{cr}^2 \frac{RSI_0 - RSI_{meas}}{RSI_{max} - RSI_0}} = N_{max} \sqrt{\frac{RSI_{meas} - RSI_{min}}{RSI_{max} - RSI_0}} \quad (13)$$

Shown in Figure 2, Equation 13 can serve as the main formula for an algorithm  $N_{app} = f(RSI_{meas})$  when the goal is to maximize the yield. The multiplier for traditionally defined  $N_{max}$  is a square root scaled by the feasible range drop of  $RSI_{meas}$  below the  $RSI_{max}$  value. This results in the multiplier equal to 0 when the  $RSI_{meas}$  value is equal to the  $RSI_{max}$  and to 1 when the  $RSI_{meas}$  equals  $RSI_0$ .

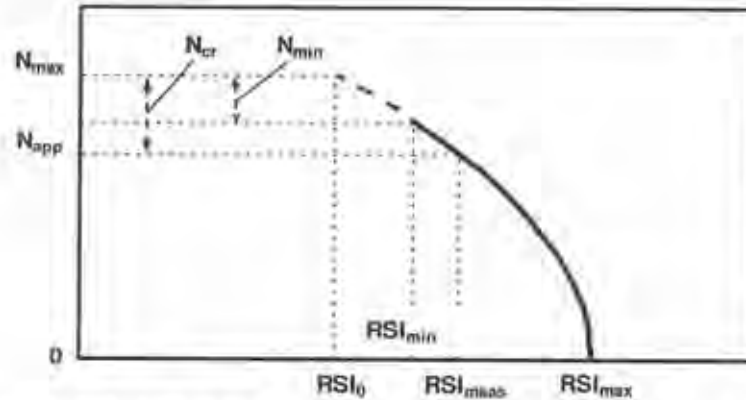


Figure 2. Sensor-based algorithm to manage in-season N to maximize the yield.

#### Profit-maximizing algorithm

With the cost of fertilizer continuing to increase, the goal of the optimization process is to maximize profit ( $P$ ), which can be defined as:

$$P = RSI_{meas} \cdot p_{RSI} - N_{app} \cdot c_N \quad (14)$$

where:

$c_N$  = monetary cost of N (same units as in Equation 3);

$p_{RSI}$  = monetary price (value) of RSI associated with predicted crop yield.

Based on a linear relationship between yield and RSI, it is safe to assume that  $p_{RSI}$  is a scaled value for the unit price of yield ( $p_Y$ ). As an illustration, this can be calculated using:

$$p_{RSI} = p_Y \frac{Y_{max} - Y_{min}}{RSI_{max} - RSI_{min}} \quad (15)$$

where:

$p_Y$  = monetary price of yield;

$Y_{min}$  and  $Y_{max}$  = minimum and maximum yield estimates that correspond to  $RSI_{min}$  (check plot) and  $RSI_{max}$  (reference plot),

Substituting Equation 10 into Equation 15 produces:

$$P = p_{RSI} \left( a_0 + a_1 (N_{cr} + N_{app}) + a_2 (N_{cr} + N_{app})^2 \right) - c_N N_{app} \quad (16)$$

To maximize the profit,  $dP/dN_{app} = 0$ , which means that:

$$\frac{dP}{dN_{app}} = a_1 p_{RSI} + 2a_2 p_{RSI} N_{cr} + 2a_2 p_{RSI} N_{app} - c_N = 0 \quad (17)$$

Solving for  $N_{app}$  again produces:

$$N_{app} = -\frac{a_1}{2a_2} + \frac{c_N}{2a_2 p_{RSI}} - N_{cr} = N_{max} + \frac{c_N}{2a_2 p_{RSI}} - N_{cr} \quad (18)$$



Defining  $N_{opt}$  as profit maximizing sum of  $N_{cr}$  and  $N_{app}$  and using  $a_2$  defined in Equation 7 produces:

$$N_{opt} = N_{max} + \frac{c_N}{2p_{RSI}a_2} = N_{max} - \frac{c_N N_{max}^2}{2p_{RSI}(RSI_{max} - RSI_0)} \quad (19)$$

Using Equations 9 and 19, Equation 18 can be rewritten as:

$$N_{app} = N_{max} - \frac{c_N N_{max}^2}{2p_{RSI}(RSI_{max} - RSI_0)} - N_{max} + \sqrt{N_{max}^2 - \frac{RSI_0 - RSI_{meas}}{a_2}} \quad (20)$$

This may be rewritten in a more convenient form:

$$N_{app} = N_{max} \left( \sqrt{\frac{RSI_{max} - RSI_{meas}}{RSI_{max} - RSI_0}} - \frac{N_{max}}{2(RSI_{max} - RSI_0)} \cdot \frac{c_N}{p_{RSI}} \right) \quad (21)$$

The N application rate obtained from economic optimization Equation 21 approaches the yield optimization Equation 13 when the ratio  $c_N/p_{RSI}$  is small. As an illustration of this, Figure 3 shows Equation 14 with three different  $c_N/p_{RSI}$  ratios. The resulting, recommended algorithms (Equation 21) are shown in Figure 4. The variables used for these illustrations are summarized in Table 1. The corn response equation is similar to the one presented by Solari *et al.* (2010). For comparison, a stronger response equation was used (two times greater  $RSI_{max} - RSI_0$ ) with two  $c_N/p_{RSI}$  ratios.  $p_{RSI}$  value was estimated according to Equation 15.

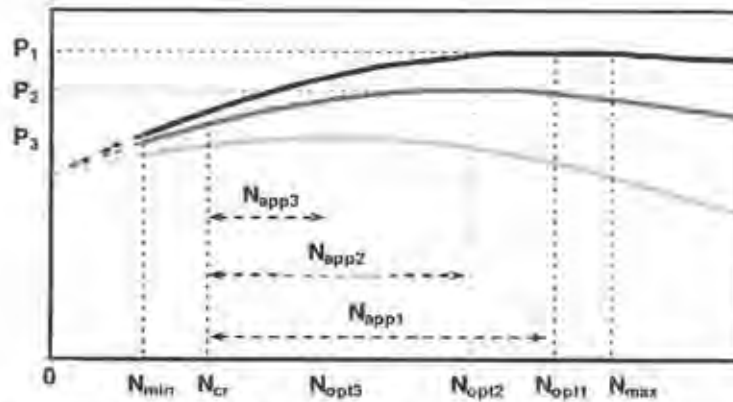


Figure 3. Profit response function for three different  $c_N/p_{RSI}$  ratios.

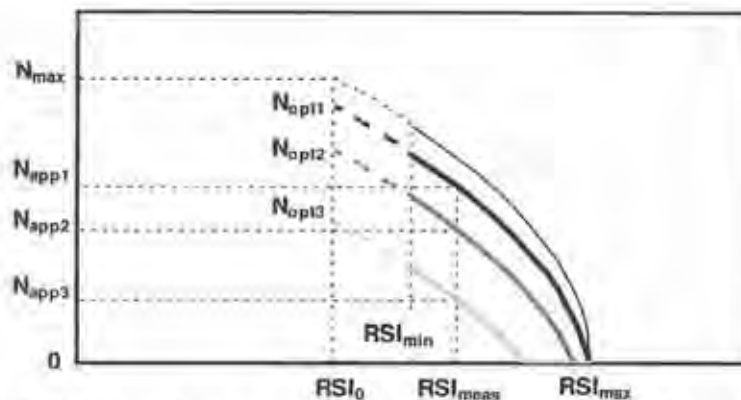


Figure 4. Sensor-based algorithm to manage in-season N to maximize the profit assuming three different  $c_N/p_{RSI}$  ratios.

Table 1. Summary of different scenarios for corn used for Figure 3 and 4.

Response	Moderate ( $RSI_0=0.8$ )				Strong ( $RSI_0=0.6$ )	
	Max yield <sup>1</sup>	Case 1	Case 2	Case 3	Max yield	Case 3
$N_{max}$ , kg/ha	180	180	180	180	180	180
$RSI_0$	0.8	0.8	0.8	0.8	0.6	0.6
$RSI_{max}$	1	1	1	1	1	1
$a_0$	0.8	0.8	0.8	0.8	0.6	0.6
$a_1$	0.0022	0.0022	0.0022	0.0022	0.0044	0.0044
$a_2$	-6.17E-06	-6.17E-06	-6.17E-06	-6.17E-06	-1.23E-05	-1.23E-05
$N_{min}$	30	30	30	30	30	30
$RSI_{min}$	0.86	0.86	0.86	0.86	0.72	0.72
$Y_{min}$ , Mg/ha	12.1	12.1	12.1	12.1	10.1	10.1
$Y_{max}$ , Mg/ha	14	14	14	14	14	14
$c_N$ , \$/kg	0	1	2	3	0	3
$P_{yield}$ , \$/Mg	200	200	200	200	200	200
$P_{RSI}$ , \$/ha	1,944	1,944	1,944	1,944	1,944	1,944
$c_N/P_{RSI}$	0.00000	0.00051	0.00103	0.00154	0.00000	0.00154
$N_{opt}$ , kg/ha	180	138	97	55	180	118
$N_{crit}$ , kg/ha	50	50	50	50	50	50
$RSI_{meas}$	0.90	0.90	0.90	0.90	0.79	0.79
$N_{app}$ , kg/ha	130	88	47	5	130	68

<sup>1</sup> Italic font indicates assumed parameters.

## Discussion

The illustration shows that an increase in the ratio  $c_N/P_{RSI}$  may result in a significant reduction in the recommended N application rate. The last column indicates that this reduction is lower for situations when there is a stronger crop response to N fertilization. However, it is independent from  $RSI_{meas}$  and is constant for a fixed N response function ( $a_0$ ,  $a_1$ ,  $a_2$ , or  $N_{max}$ ,  $RSI_0$ ,  $RSI_{max}$ ).

While analyzing Equation 21 and its partial form, Equation 13, it is obvious that the original assumption of second-order polynomial crop response to plant available N results in the recommended algorithm with a rapid increase of  $N_{app}$  as the  $RSI_{meas}$  value falls slightly behind  $RSI_{max}$  ( $RSI_{opt}$ ). The rate of this increase is becoming lower as higher application rates are required. The problem is that spatial variability of soil water supply and the nutrient storage capacity makes the definition of  $RSI_0$  and  $RSI_{max}$  (inherently  $RSI_{opt}$ ) inconsistent across the field. High-density thematic soil maps (topography, water supply potential, organic nitrogen mineralization potential, etc.) can be used to partially account for this spatial variability. Yet, the relatively high change in  $N_{app}$  when  $RSI_{meas}$  approaches  $RSI_{opt}$  amplifies the effect of relevant inconsistencies.

Furthermore, it is important to remember that the derived equations rely on a linear relationship between yield and RSI for the entire range of N. It assumes that both variables follow the same response to N availability with no difference between prior N availability ( $N_{av}$ ) and N applied during crop canopy sensing ( $N_{app}$ ). This is definitely not valid due to the potential for non-recoverable crop stress and temporary inconsistent N uptake. In other words, a poor performing crop will not necessarily improve with plenty of in-season fertilization and a healthy crop will not necessarily stay healthy until the end of growing season without supplemental fertilization.

To overcome these limitations, supplemental information can be used to adjust the algorithm. For example, Holland and Schepers (2010, 2011) implemented an exponential adjustment of an equivalent to Equation 13, with two added parameters (threshold and back-off). This facilitates a smooth decrease in the recommended N application rate in response to non-recoverable crop yield loss.

Despite the shape of the N application recommendation, it is important to note that an increase in the  $c_N/p_{RSI}$  ratio leads to a reduction of  $N_{app}$ . This means that for the same value of  $N_{cr}$ , the targeted N availability for the crop is getting smaller ( $N_{opt} < N_{max}$ ). However, it is important to remember that both crop response and marketing variables have significant uncertainties, which in practice lead to more conservative strategies. For example, this could be achieved by setting up a positive minimum  $N_{app}$  for high values of  $RSI_{meas}$ . In any case, both economic and risk management considerations have to be taken into account when developing strategies for crop canopy sensing-based in-season N management in different crop scenarios and in various climatic regions. This will be considered in a follow-up study.

## Conclusions

A theoretical exploration of a popular, in-season, nitrogen management technique that relies on the second-order polynomial response function illustrated that the recommended algorithm is sensitive to the site-specific scaling parameters of a relative sensor index. Economic considerations suggest a reduction in the nitrogen application rate when the cost of fertilizer increases relative to crop price. This reduction is greater when crop response to N fertilization is not strong and smaller when N deficiency could substantially reduce the relative sensor index and the ultimate yield. However, a risk management approach suggests that a more involved mathematical methodology could be pursued to address the uncertainties associated with establishing the critical values of the relative sensor index. The derived equations are influenced by the accepted assumptions and should be re-assessed when developing new algorithms for practical crop canopy sensor-based variable rate nitrogen application to maximize crop production profitability.

## References

- Biggs, G.L., Blackmer, T.M., Demetriades-Shah, T.H., Holland, K.H., Schepers, J.S. and Wurm, J.H. 2002. Method and apparatus for real-time determination and application of nitrogen fertilizer using rapid non-destructive crop canopy measurements. U.S. Patent no. 6,393,927. Date issued: May 28, 2002.
- Bongiovanni, R. and Lowenberg-DeBoer, J. 2000. Economics of variable rate lime in Indiana. *Precision Agriculture* 2 55-70.
- Bullock, D. and Bullock, D. 1994. Calculation of optimal nitrogen fertilization rates. *Agronomy Journal* 86 921-923.
- Havlin, J.L., Tisdale, L. H., Nelson, W.L. and Beaton, J.D. 2004. *Soil Fertility and Fertilizers: An Introduction to Nutrient Management*, Seventh Edition. Prentice Hall, Upper Saddle River, New Jersey, USA.
- Holland, K.H. and Schepers, J.S. 2010. Derivation of a variable rate nitrogen application model for in-season fertilization of corn. *Agronomy Journal* 102 (5) 1415-1424.
- Holland, K. and Schepers, J. S. 2011. Use of a virtual-reference concept to interpret active crop canopy sensor data. *Precision Agriculture* 14 (1) 71-85.
- Kitchen, N.R., Sudduth, K.A., Drummond, S.T., Scharf, P.C., Palm, H.L., Roberts, D.F. and Vories, E.D. 2010. Ground-based canopy reflectance sensing for variable-rate nitrogen corn fertilization. *Agronomy Journal* 102 71-84.
- Samborski, S.M., Tremblay, N. and Fallon, E. 2009. Strategies to make use of plant sensors-based diagnostic information for nitrogen recommendations. *Agronomy Journal* 101 800-816.



- Shiratsuchi, L., Ferguson, R., Shanahan, J., Adamchuk, V., Rundquist, D., Marx, D. and Slater, G. 2011. Water and nitrogen effects on active canopy sensor vegetation indices. *Agronomy Journal* 103(6) 1815-1826.
- Solari, E., Shanahan, J., Ferguson, R., Schepers, J., Gitelson, A. 2008. Active sensor reflectance measurements of corn nitrogen status and yield potential. *Agronomy Journal* 100 571-579.
- Solari, E., Shanahan, J.F., Ferguson, R.B. and Adamchuk, V. I. 2010. An active sensor algorithm for corn nitrogen recommendations based on a chlorophyll meter algorithm. *Agronomy Journal* 102(4): 1090-1098.

# Effect of sampling patterns and interpolation methods on prediction quality of soil variability mapping

H.H. Huang<sup>1</sup>, V.I. Adamchuk<sup>1</sup>, I.I. Boiko<sup>2</sup> and R.F. Ferguson<sup>3</sup>

<sup>1</sup>McGill University, Ste-Anne-de-Bellevue, QC, H9X 3V9, Canada; [viacheslav.adamchuk@mcgill.ca](mailto:viacheslav.adamchuk@mcgill.ca)

<sup>2</sup>Druzhiba-Nova, LLC, Varva, Cherkiv Reg., Ukraine

<sup>3</sup>University of Nebraska-Lincoln, Lincoln, NE 68583, USA

## Abstract

Precision agriculture technologies have benefited from a number of remote and proximal sensing mapping practices. However, systematic (grid) sampling of soil still remains the most popular method to address spatial heterogeneity for an array of soil chemical properties. Because of economic constraints, the density of grid sampling has been limited. To increase the information value of the data obtained, at an affordable cost, different sampling and data interpolation strategies have been used. This paper examines several point-based and composite sampling methods implemented using square grids by combining multiple soil cores using a semi-automatic soil sampling systems.

**Keywords:** soil sampling, data interpolation, thematic soil maps

## Introduction

Knowledge of the spatial variability of the chemical and physical properties of soil is essential for nutrient application for both conventional fertilization and when using variable-rate technology. Although proximal soil sensing technology (Viscarra Rossel *et al.*, 2011) has been used to obtain high-density data, revealing the spatial distribution of influential soil attributes across an agricultural field, systematic soil sampling according to a predefined grid pattern (Wollenhaupt *et al.*, 1997) remains the most popular practice to map the spatial variability of the indicators for soil fertilization. According to Brouder *et al.* (2005), the accuracy of soil variability mapping relies on an adequate intensity of sampling and the proper prediction (interpolation) method. Kerry and Oliver (2004) suggested that to ensure the quality of soil variability prediction, a variogram can be used to determine the optimal sampling interval. They recommended that sampling should occur at an interval of less than half the range of spatial dependence or to determine the acceptable error by incorporating the variogram into the kriging interpolation.

Although the soil sampling and data interpolation process should be optimized for a specific soil attribute and sampling site (Van Groenigen and Stein, 1998; Webster and Oliver, 2006; Brus and Heuvelink, 2007; McBratney and Pringle, 2009; Buczko *et al.*, 2012), precision agriculture service providers tend to implement strategies consistent through their entire operation. Thus, Gotway *et al.* (1996), Shi *et al.* (2000), Mueller *et al.* (2001) and Brouder *et al.* (2005) provided examples of research projects aimed at a comparison of alternative soil sampling and data analysis strategies. In general, it was noted that larger numbers of independently analyzed soil samples have a better chance of revealing spatial heterogeneity of soil fertility indicators. The difference between alternative data interpolation methods was frequently insignificant and similar to the field averaging method. Systematic soil sampling is achieved by either collecting a given number of fixed depth soil cores from a single location within each grid (point sampling), or by collecting these cores from the entire grid cell (area sampling). Although many practitioners prefer point sampling because of the time involved, the recent availability of semi-automated soil sampling systems and easy-to-use navigation terminals (Figure 1) make point and area sampling strategies comparable in terms of the required time and cost. Naturally this poses a question whether a specific grid-based soil sampling



Figure 1. Semi-automated Nietfeld 2005 soil sampling system (left) with in-cab navigation (right) used in this research.

can provide better representation of the spatial distribution of selected soil attributes once the respective interpolated surfaces are constructed. Ultimately, implementing precision agriculture on a large scale requires a methodology that could be robust from field to field, yet sensitive to substantial spatial soil heterogeneity.

Among an infinite number of possibilities, the objective of this study was to evaluate the prediction quality of selected soil properties using three types of grid-based sampling patterns (center-point-grid-sampling, diagonal-grid-cell-sampling and z-pattern-grid-cell-sampling) while obtaining maps through common interpolation methods: ordinary kriging, inverse distance weighting (IDW), and nearest neighbour (tiling).

### Materials and methods

In 2012, two agricultural fields (Field 1 and Field 2) located in north-central Ukraine were chosen to collect soil samples using three different 1 ha square grid design patterns (Figure 2). Both fields were similar in area, approximately 47 ha. This resulted in 46 grid cells in Field 1 and 47 grid cells in Field 2. Grid cells touching the boundary were adjusted to assure complete field coverage; they could be irregular in shape as well as somewhat smaller or larger than 1 ha. Composite soil samples were obtained as: (1) 15 soil cores from a 5×5 m area around the center of each grid cell (center-point-grid-sampling); (2) 20 soil cores collected from one of two diagonals for each grid cell (diagonal-grid-cell-sampling); and (3) 20 soil cores obtained while making a z-pattern grid within each grid cell (z-pattern-grid-cell-sampling). In addition to grid samples, 12 validation samples for Field 1 and 20 validation samples for Field 2 were collected at randomly scattered locations in a similar way to center-point-grid sampling.

All of these composite samples were analyzed in laboratory conditions according to the State Committee of Ukraine for Standardization (DSTU, 2005). Soil variability predictions were performed on phosphorus ( $P_2O_5$ ) and potassium ( $K_2O$ ) for Field 1, and phosphorus ( $P_2O_5$ ), potassium ( $K_2O$ ) and total mineral nitrogen ( $NH_4 + NO_3$ ) for Field 2. Maps representing each soil property and sampling strategy were obtained by: (1) averaging results from all grid cells; (2) ordinary kriging interpolation with a 5×5 m raster resolution; (3) IDW interpolation with the same resolution; and (4) tiling when the entire grid cell was assigned the value of a corresponding test report. Root mean squared error (RMSE) was calculated for each combination of soil sampling and data interpolation using validation samples. The ratio of MSE (squared RMSE) was used to check the significance of the difference in soil property prediction quality at  $\alpha=0.05$ .



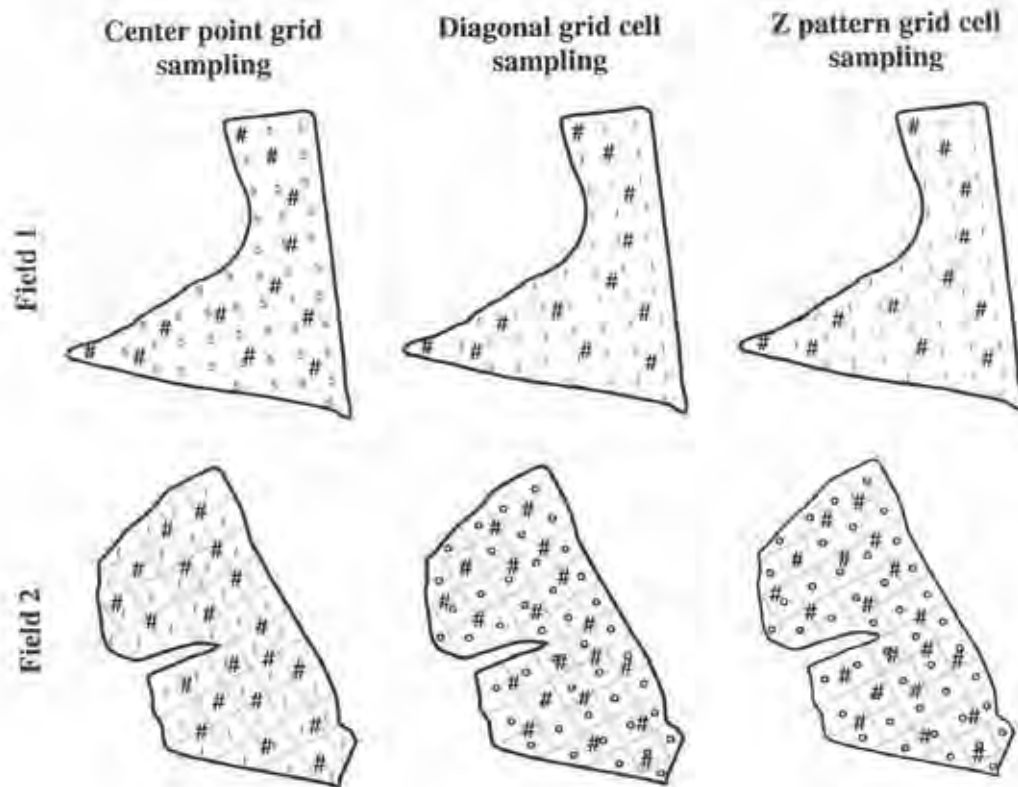


Figure 2. Three sampling patterns evaluated in the project.

To strengthen the comparison, apparent soil electrical conductivity  $EC_a$  data from six production fields in Nebraska (BR, HE, KR, RA, HU, and LU) were used to model the same sampling strategies. These fields represent different levels of spatial structure and were obtained using a Veris® 3100 EC Surveyor with 3-12 m intervals between passes. Field areas ranged from 25 to 60 ha. Dense  $EC_a$  sensor-based point measurements were transformed into a raster with 5×5 m pixels, and then converted into corresponding vector point data. Then, three sampling strategies and four interpolation procedures similar to those obtained in the Ukrainian fields were applied with both 1 ha and 5 ha size grids. Due to the different field sizes, the number of grid cells per field ranged from 27 to 60 for the 1 ha grid and from 6 to 16 for the 5 ha grid. Unlike chemical soil properties,  $EC_a$  data allowed for the estimation of RMSE values using 5×5 m raster pixels, thus, allowing for a greater number of validation points (between 6,712 and 13,9587) when comparing  $EC_a$  prediction quality using different modeled sampling and data interpolation strategies.

### Result and discussion

Figure 3 illustrates the semi-variograms for the two Ukrainian fields. They were calculated using center-point-grid and validation samples together. A spherical model was applied in each case for consistency as other models did not indicate significant benefits. Although none of these semi-variograms illustrated a strong spatial structure, the distributions of P in Field 1 and K in Field 2 revealed some reduction in the difference between the values corresponding to points close to each other.

When comparing the RMSE for different sampling/interpolation combinations (Figure 4), it could be seen that the z-pattern method, with either kriging or IDW interpolation, resulted in RMSE values among the lowest in each case. Since only 12 validation samples were obtained in Field 1, all observed

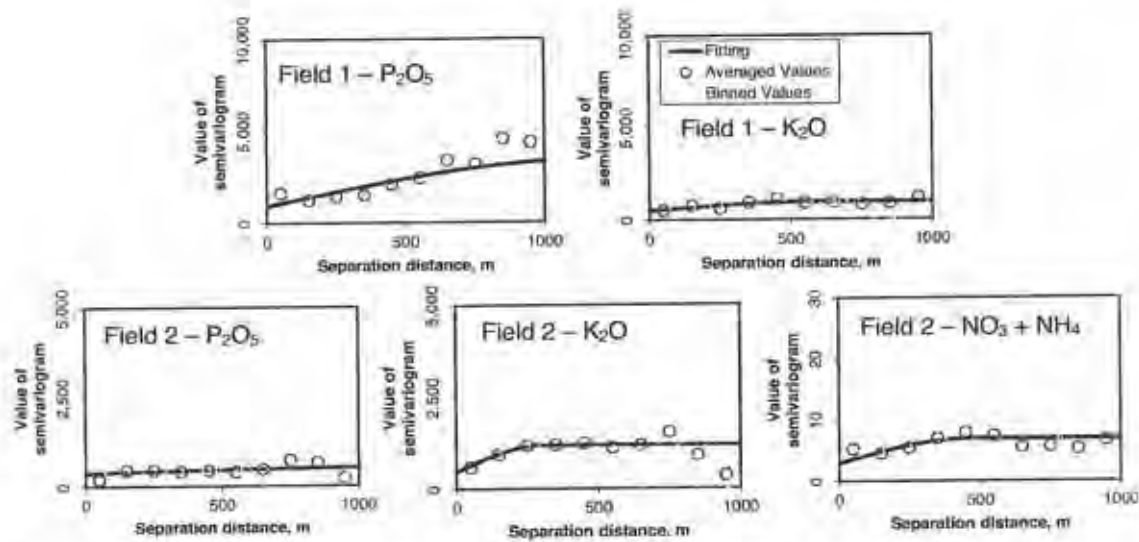


Figure 3. Semi-variograms of the chemical properties for two Ukrainian fields.

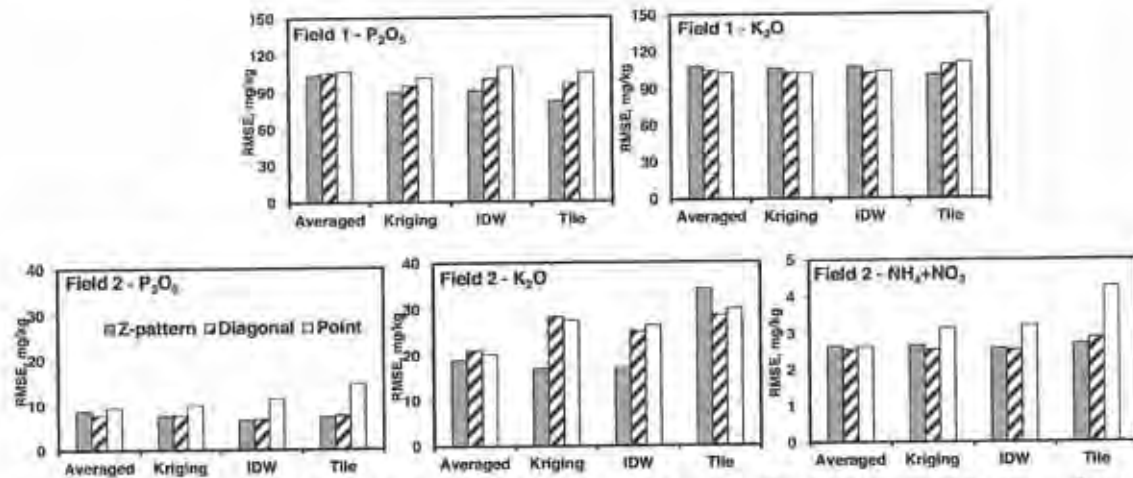


Figure 4. Comparison of soil chemical property prediction quality for the two Ukrainian fields.

differences were not significant at  $\alpha=0.05$ . In Field 2, tiling of center-point-grid measurements of mineral nitrogen produced RMSE values significantly higher than the area sampling methods. A similar observation was found for phosphorous. For potassium, z-pattern-grid-cell sampling with kriging and IDW interpolation produced error estimates significantly lower than some alternatives. However, none of these cases produced field averaging RMSE values significantly different from interpolated maps.

Figure 5 illustrates the semi-variograms for soil ECa in the six Nebraska fields. Due to the complex spatial structure, two nested spherical semi-variogram models were fitted in each case. There was no significant difference between the spherical and the alternative models. It appeared that field BR was relatively uniform, fields HE, KR, and LU had spatial structures with comparable range values for both nested semi-variograms, and fields RA and HU indicated less than a 100 m range for short separation distance variability and longer ranges for greater distance variability.

When comparing RMSE values (Figure 6), one could see once again a general reduction in RMSE when switching from center-point-grid sampling to diagonal-grid-cell sampling and further to

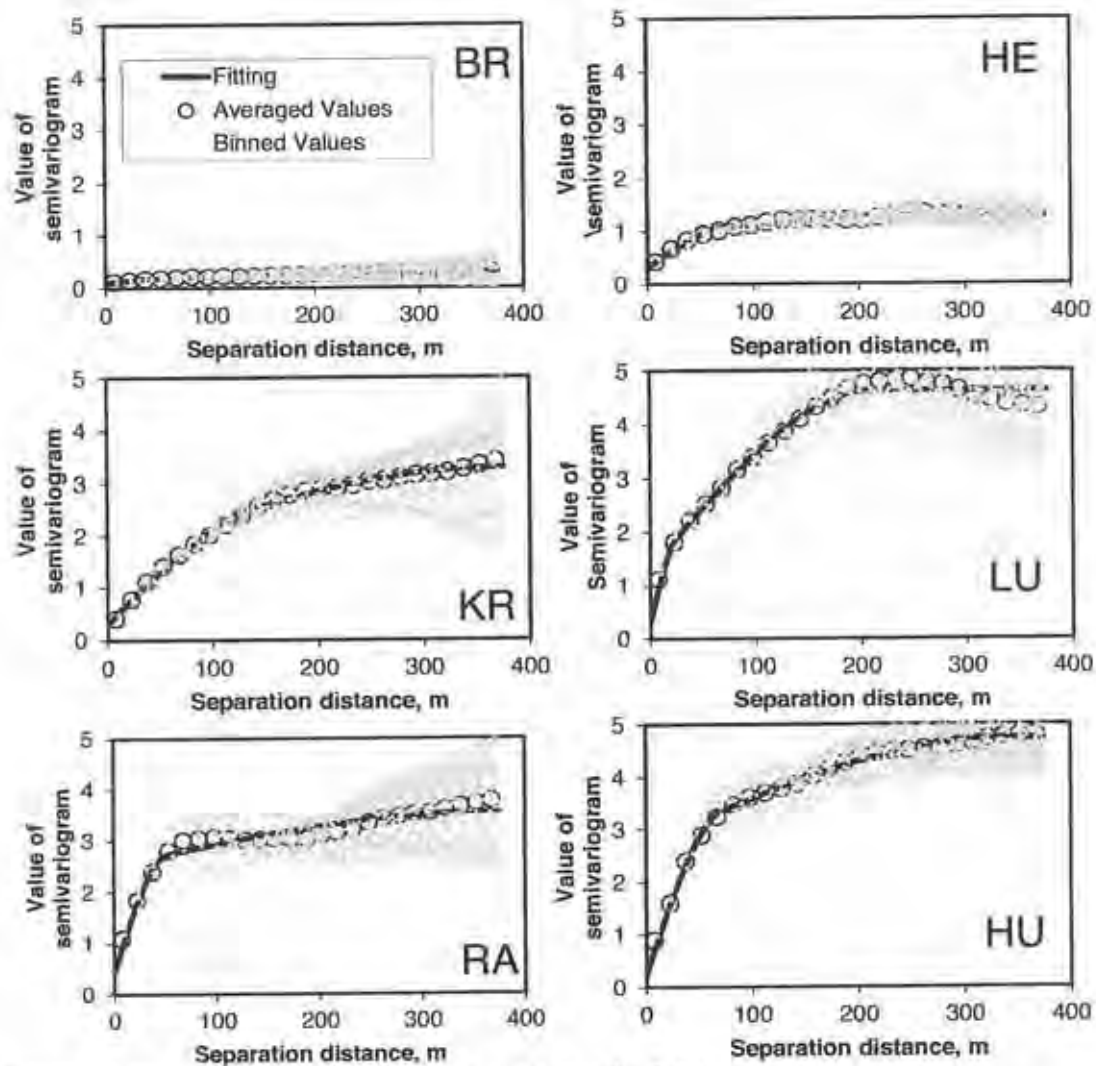


Figure 5. Semi-variograms of soil  $EC_e$  for six Nebraska fields.

z-pattern-grid-cell sampling as well as when moving from field averaging and tiling to kriging and IDW interpolation. However, these observations did not repeat in each instance. Due to a large number of raster pixels available for evaluation, even slight differences were statistically significant. Yet, in comparing the results, test conclusions were different for each field.

It was obvious that the 5 ha grid sampling always produced RMSE values higher than the corresponding 1 ha grid sampling (except for the field average approach). However, it was interesting to note that the most noticeable RMSE increases for center-point-grid sampling occurred when using 5 ha sampling. This could be attributed to specific grid cells with relatively high levels of heterogeneity; the aggregating measurements from around those grids were more conforming than relying on the estimate from a single location.

Similar to previous studies, these results failed to isolate a single soil sampling and data interpolation strategy that would be favourable in each case. Furthermore, 1 ha grid sampling could not produce RMSE values consistently lower than the field average approach. However, it was noted that the interpolated maps based on z-pattern-grid-cell sampling did not produce RMSE estimates significantly higher than the alternatives throughout the entire study. This suggests that a smoothing effect due to the area rather than point sampling can reduce the undesired misrepresentation of



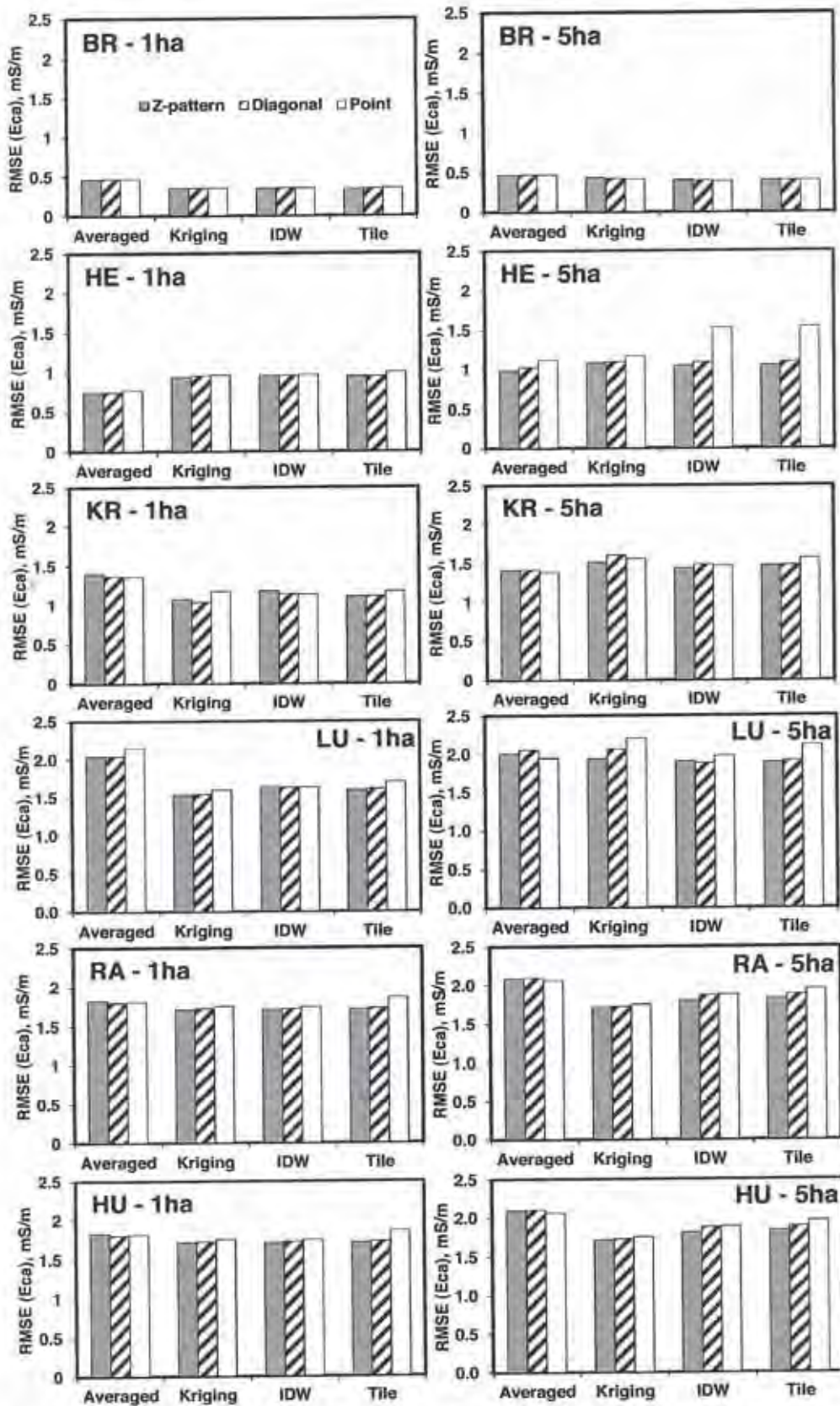


Figure 6. Comparison of soil EC<sub>a</sub> prediction quality for the six Nebraska fields.

certain parts of each grid. Thus area sampling could produce maps similar, if not better, in quality to the field average maps that could be generated by averaging all grid-based measurements (different from traditional 20 ha or greater sampling strategies). A reduction in error estimates when using systematic soil sampling was apparent with a stronger spatial structure of the soil property being measured. Further studies related to this research could help in determining the best soil sampling and data processing techniques that would take full advantage of semi-automated soil sampling equipment. On-going research on adaptive soil sampling strategies that take into account high-resolution proximal and remote sensing data related to soil fertility indicators is a promising alternative to systematic grid-based sampling, which is still pursued commercially as the main practice.

## Conclusion

In this study, three types of grid-based systematic soil sampling and different data interpolation approaches were compared. Two production fields were sampled using three methods to construct maps of soil chemical properties; sensor-based  $EC_e$  measurements from six production fields were used to model the same sampling and data interpolation strategies. It was obvious that a larger sampling grid reduced the potential to uncover low-range variability. Although none of the alternative methods produced significantly smaller RMSE estimates throughout the entire study, in many cases, z-pattern-grid-cell sampling, with either ordinary kriging or IDW interpolation, indicated a tendency to provide lower prediction errors than other types of sampling or tiling, when the entire grid cell was assigned the value corresponding to the composite soil sample obtained within that grid. On the other hand, point-based sampling with tiling resulted in a few cases of the weakest predictions. It has also been noted that the presence of a spatial structure was an important ingredient for an interpolated thematic soil map that results in more accurate information than the field average map. Further research may help to establish a robust method of soil sampling and data processing that provides consistently more accurate thematic soil maps in fields with substantial spatial structure and predictions similar to the field average approach in cases where change of a given soil attribute is negligible or stochastic through the entire field.

## References

- Brouder, S.M., B.S. Hofmann and D.K. Morris. 2005. Mapping Soil pH. *Soil Science Society of America Journal* 69(2) 427-442.
- Brus, D.J., and G. Heuvelink 2007. Optimization of sample patterns for universal kriging of environmental variables. *Geoderma* 138(1) 86-95.
- Buczko, U., U. Kurfürst, R.O. Kuchenbuch and M. Munzert. 2012. Factors affecting sampling error of nutrient element contents of arable fields. *Communications in Soil Science and Plant Analysis* 43(11) 1576-1598.
- DSTU, 2005. Quality of soil. Measuring mobile compounds of phosphorous and potassium using Kirsanov method modified by NNTIGA (in Ukrainian: Yakist' gruntu. Vyznachennia ruhomyh spoluk fosforu i kaliyu za metodom Kirsanova v modyfikatsii NNTIGA). DSTU (State Committee of Ukraine for Standardization) 4405.
- Gotway, C.A., R.B. Ferguson, G.W. Hergert and T.A. Peterson. 1996. Comparison of kriging and inverse-distance methods for mapping soil parameters. *Soil Science Society of America Journal* 60 1237-1247.
- Kerry, R. and M.A. Oliver. 2004. Average variograms to guide soil sampling. *International Journal of Applied Earth Observation and Geoinformation* 5(4) 307-325.
- McBratney, A.B. and M.J. Pringle. 1999. Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. *Precision Agriculture* 1(2) 125-152.
- Mueller, T.G., F.J. Pierce, O. Schabenberger and D.D. Warncke. 2001. Map quality for site-specific fertility management. *Soil Science Society of America Journal* 65(5) 1547-1558.

- Shi, Z., K. Wang, J.S. Bailey, C. Jordan and A.J. Higgins. 2000. Sampling strategies for mapping soil phosphorus and soil potassium distributions in cool temperate grassland. *Precision Agriculture* 2(4) 347-357.
- Van Groenigen, J.W. and A. Stein. 1998. Constrained optimization of spatial sampling using continuous simulated annealing. *Journal of Environmental Quality* 27(5) 1078-1086.
- Viscarra Rosset, R.A., V.I. Adamchuk, K.A. Sudduth, N.J. McKenzie and C. Lobsey. 2011. Proximal soil sensing: an effective approach for soil measurements in space and time, Chapter 5. *Advances in Agronomy* 113 237-283.
- Webster, R. and M.A. Oliver. 2006. Sample adequately to estimate variograms of soil properties. *Journal of Soil Science* 43(1) 177-192.
- Wollenhaupt, N.C., D.J. Mulla and C.A. Gotway Crawford. 1997. Soil sampling and interpolation techniques for mapping spatial variability of soil properties, p. 19-53. In: E.J. Pierce and E.J. Sadler (ed.) *The State of Site-Specific Management for Agriculture*. ASA-CSSA-SSA, Madison, Wisconsin, USA.



Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

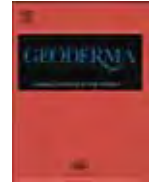
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/authorsrights>



Contents lists available at SciVerse ScienceDirect

Geoderma

journal homepage: [www.elsevier.com/locate/geoderma](http://www.elsevier.com/locate/geoderma)

## Preface

## Special issue on proximal soil sensing

This special issue presents a collection of articles describing the research presented at the Second Global Workshop on Proximal Soil Sensing (PSS) hosted by McGill University, Montreal, Quebec, Canada on May 15–18, 2011 and held under the auspices of the International Union of Soil Sciences (IUSS) Working Group on Proximal Soil Sensing (WG-PSS). Sixty scientists from 18 countries participated in the workshop where current scientific and technology challenges pertaining to the field of PSS were discussed.

PSS is defined as the use of field-based sensors to obtain signals from the soil when the sensors' detector is in contact with, or close proximity to (within 2 m) the soil (Viscarra Rossel and McBratney (1998), Viscarra Rossel et al., 2011). The sensors detect signals corresponding to physical measures, which can be linked to different soils and their properties. According to a proposed classification (Viscarra Rossel et al., 2011), proximal soil sensors may be described by the manner in which they measure (invasive or non-invasive), the source of their energy (active or passive), the way in which they operate (stationary or mobile), and the inference used in measuring the target soil property (direct or indirect).

In terms of the physical phenomena involved in the measurements, a large number of PSS systems use the soil's ability to reflect or emit energy in different parts of the electromagnetic spectrum from high-frequency gamma-rays and X-rays to ultraviolet, visible, infra-red and radio-waves. PSS systems may rely on the ability of soil particles to conduct and accumulate an electrical charge, while others quantify the mechanical interaction between the sensor and soil, and others use ion-selective potentiometry.

Whatever the approach, proximal soil sensors facilitate the collection of large amounts of (spatial and temporal) data using cheaper, simpler, and less laborious techniques, which, as a whole, are very informative. The measurements can be made in the field from the surface or within the soil profile, and this information is produced almost instantaneously. Hummel et al. (1996), Sudduth et al. (1997), Adamchuk et al. (2004) and Shibusawa (2006) provide a review of mobile (on-the-go) proximal soil sensing solutions. A book entitled "Proximal Soil Sensing" (Viscarra Rossel et al., 2010) includes selected research articles presented at the first High Resolution Digital Soil Sensing and Mapping Workshop hosted by the University of Sydney, Sydney, Australia on February 5–8, 2010.

This special issue begins with the work of Lueck and Ruehlmann as well as that of Sudduth et al. who explore the extendibility of electrical resistivity mapping to determine both spatial and vertical changes in soil physical characteristics. Hedley et al. illustrate the potential for the use soil electrical conductivity maps to be integrated with a wireless sensor network to optimize soil water management. Although

most of the articles pertain to agricultural landscapes, De Shmedt et al. used electromagnetic induction sensing as an archeological tool. Articles by Nocita et al., Ramirez-Lopez et al., Brodsky et al., Kodaira et al., and Kweon et al. focus on the analysis of optical soil reflectance measurements in the visible and near-infrared parts of the electromagnetic spectrum. Dierke and Werban have studied the capabilities of gamma-ray radiometry, while Van Meirvenne et al. and Piikki et al. have looked into combining gamma-ray with other sensors to improve the quality of soil measurements. De Benedetto et al. have integrated several different types of PSS systems. Finally, Coulouma et al. discuss the use of seismic sensors.

This selection of diverse and interesting research results will extend your knowledge of the applicability of PSS methods to improve your understanding of spatial and temporal soil variability. We hope that this issue will stimulate additional research worldwide to examine new and creative approaches to proximal soil sensing. We also invite any interested colleagues to participate in the future activities coordinated through the IUSS WG-PSS ([www.proximalsoilsensing.org](http://www.proximalsoilsensing.org)).

## References

- Adamchuk, V.I., Hummel, J.W., Morgan, M.T., Upadhyaya, S.K., 2004. On-the-go soil sensors for precision agriculture. *Computers and Electronics in Agriculture* 44, 71–91.
- Hummel, J.W., Gaultney, L.D., Sudduth, K.A., 1996. Soil property sensing for site-specific crop management. *Computers and Electronics in Agriculture* 14, 121–136.
- Shibusawa, S., 2006. Soil sensors for precision agriculture. In: Srinivasan, A. (Ed.), *Handbook of Precision Agriculture. Principles and Applications*. Food Products Press, New York, New York, USA, pp. 57–90.
- Sudduth, K.A., Hummel, J.W., Birrell, S.J., 1997. Sensors for site-specific management. In: Pierce, F.T., Sadler, E.J. (Eds.), *The State of Site-Specific Management for Agriculture*. ASA-CSSA-SSSA, Madison, Wisconsin, USA, pp. 183–210.
- Viscarra Rossel, R.A., McBratney, A.B., 1998. Laboratory evaluation of a proximal sensing technique for simultaneous measurement of clay and water content. *Geoderma* 85, 19–39.
- Viscarra Rossel, R.A., McBratney, A., Minasny, B. (Eds.), 2010. *Proximal Soil Sensing*. Springer, New York, New York, USA.
- Viscarra Rossel, R.A., Adamchuk, V.I., Sudduth, K.A., McKenzie, N.J., Lobsey, C., 2011. Proximal soil sensing: an effective approach for soil measurements in space and time, chapter 5. *Advances in Agronomy* 113, 237–283.

Viahceslav Adamchuk  
Guest Editor

Raphael A. Viscarra Rossel  
Guest Editor

# *Analysis of soil water availability by integrating spatial and temporal sensor- based data*

**L. Pan, V. I. Adamchuk, D. L. Martin,  
M. A. Schroeder & R. B. Ferguson**

## **Precision Agriculture**

An International Journal on Advances in  
Precision Agriculture

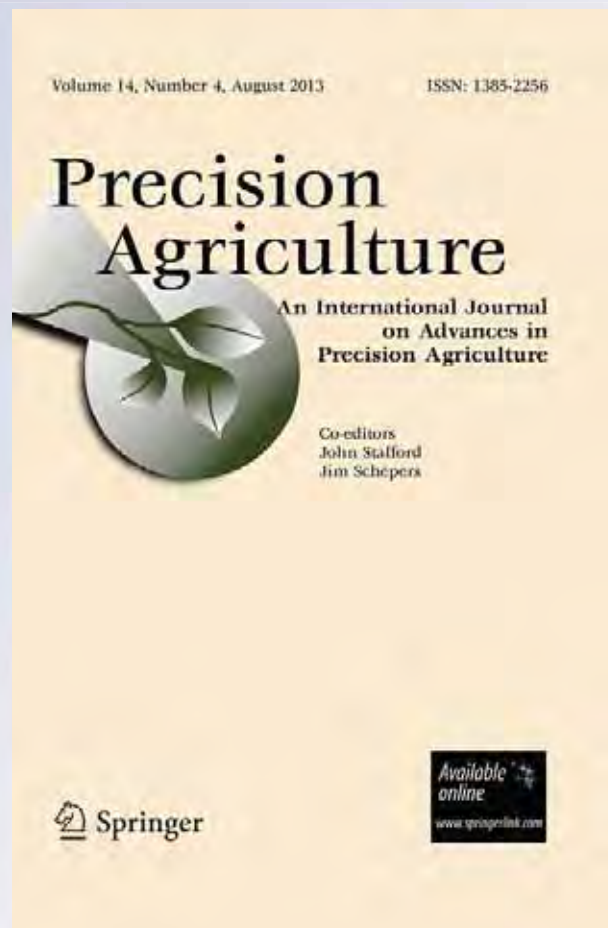
ISSN 1385-2256

Volume 14

Number 4

Precision Agric (2013) 14:414-433

DOI 10.1007/s11119-013-9305-x





**Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media New York. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**

## Analysis of soil water availability by integrating spatial and temporal sensor-based data

L. Pan · V. I. Adamchuk · D. L. Martin · M. A. Schroeder ·  
R. B. Ferguson

Published online: 7 February 2013  
© Springer Science+Business Media New York 2013

**Abstract** An efficient irrigation system should meet crop demands for water. A limited water supply may result in reductions in yield, while excess irrigation is a waste of resources. To investigate water availability throughout the growing season, on-the-go sensing technologies (field elevation and apparent electrical conductivity) were used to analyze the spatial variability of soil relevant to its water-holding capacity. High-density data layers were used to identify strategic sites to monitor changes in plant-available water over time. To illustrate this approach, nine locations in a 37-ha agricultural field were selected for monitoring the soil matric potential and temperature at four depths (18, 48, 79 and 109 cm) using wireless technology. Using a linear regression approach, a field-specific model was developed that quantified plant-available water at every field location and at specific points in time. Further analysis was used to quantify the percentage of the field that undergoes a potential shortage in water supply. These results could be used to optimize irrigation scheduling and to assess the potential for variable-rate irrigation.

**Keywords** On-the-go mapping · Wireless network · Water-use efficiency

---

L. Pan · V. I. Adamchuk (✉)  
Department of Bioresource Engineering, McGill University, Ste-Anne-de-Bellevue, QC H9X 3V9,  
Canada  
e-mail: viacheslav.adamchuk@mcgill.ca

D. L. Martin  
Department of Biological Systems Engineering, University of Nebraska-Lincoln, Lincoln, NE 68583,  
USA

M. A. Schroeder  
Agricultural Research and Development Center, University of Nebraska-Lincoln, Lincoln, NE 68583,  
USA

R. B. Ferguson  
Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68583, USA

## Introduction

Shortages of water and increases in energy costs are an incentive for many producers to consider variable-rate irrigation (VRI). The spatial variability of factors including topography, soil type, water availability, landscape features and cropping systems results in differing demands for irrigated water. In response to this variability, VRI can be executed to improve water-use efficiency, thereby, increasing productivity, conserving fuel and limiting fresh water consumption. Also, VRI can reduce nutrient leaching. An increasing number of research projects have focused on implementing VRI using two key technologies: (1) newly designed sprinkler systems capable of providing spatially varying irrigation rates and (2) software and hardware for control systems to guide VRI.

For example, King and Kincaid (2004) described laboratory testing, which indicated that their variable-flow rate sprinklers could be used for VRI. In another research project, the operational equation that described the internal connection of the flow rate, rotating speed and throw distance of the variable-rate, contour-controlled sprinkler was derived using the mathematical theory of limitation and double integral (Han and Wu 2011). This provided the fundamental principles for the design of variable-rate, contour-controlled sprinklers and square-wetted area sprinklers.

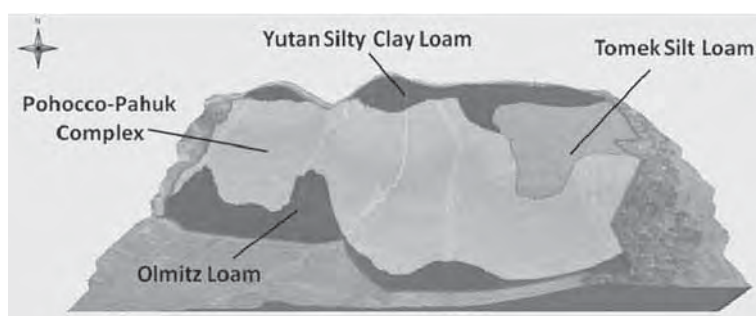
In terms of controller development, Kim and Evans (2009) reported on the development of a graphical user interface for wireless monitoring and control of sprinkler irrigators with VRI capabilities. Bao et al. (2004) integrated dielectric soil properties measurements, relevant data processing, fuzzy control arithmetic, a global positioning system (GPS) receiver and virtual instrument systems to achieve VRI as well. Hedley et al. (2009) compared VRI and uniform-rate irrigation (URI) and showed that VRI saved 9–19 % of irrigation water, reduced the loss of drainage water by 25–45 % and reduced the risk of nitrogen leaching. Moreover, the direct value of water saving using VRI was determined to be NZD \$35–\$149/ha in their study. Though the benefits of VRI are obvious, capable systems are rather expensive. However, as more producers adopt VRI systems, the cost of this technology will likely be reduced.

In addition to obtaining the key technologies aimed at the direct implementation of VRI, it is important to focus on determining the amount of the demand for irrigated water across the entire agricultural field. Conventional on-the-spot evaluation of soil conditions is somewhat subjective and labor-intensive. In addition, spatial soil variability within a field can be quite extensive, which means that defining representative locations to make irrigation management decisions is challenging.

By contrast, on-the-go soil sensing can provide fine-resolution maps of soil properties at a relatively low cost (Adamchuk et al. 2004). Such instruments provide the ability to map soil topography, apparent soil electrical conductivity ( $EC_a$ ), optical reflectance, mechanical resistance, capacitance and other properties. Unfortunately, the relationships between the data that can be collected on-the-go and agronomic soil parameters, such as water content, are frequently site-specific. In addition, the amount of water stored in the soil profile changes not only spatially, but also temporally. Sensor-based maps have been used to define the spatial variability of soil properties influencing water movement across a landscape and this information has been used to define relatively homogeneous management zones that were evaluated separately (Hedley and Yule 2009).

Wireless technology has been used increasingly to achieve temporal monitoring of soil conditions. Such systems allow the producer to obtain information about soil water status, temperature and other properties in real time from a remote location (Kim et al. 2009). Vellidis et al. (2008) developed a wireless sensor network for scheduling irrigation in





**Fig. 1** Field 1.14 at the University of Nebraska-Lincoln Agricultural Research and Development Center (Meade, Nebraska)

cotton. Radio frequency ID (RFID) tags were used to transmit the real-time soil moisture sensors' readings. The relatively low cost of the system (USD \$2400 for a 20-sensor node system) allows for the installation of a dense population of soil moisture sensors that adequately represent the inherent soil variability present in fields. This technology greatly improves the convenience of monitoring soil water status for the producer. Irrigation systems managers have used the data to optimize the use of resources in response to dynamic changes in soil conditions and to reduce the risk of crop water stress (Omary et al. 1997; Miranda et al. 2003; Han et al. 2009).

The objective of this research was to explore the methodology of soil water availability analysis through temporal monitoring of soil matric potential using a wireless sensor network based on fine-resolution maps of  $EC_a$  and field elevation.

## Materials and methods

A 37 ha field (Field 1.14) at the Agricultural Research and Development Center (ARDC) near Meade, Nebraska, USA was chosen to explore the potential to optimize water use during an irrigation season (Fig. 1). The field was mapped using a Veris<sup>®</sup> 3150 unit (Mobile Sensor Platform, Veris Technologies, Inc., Salina, Kansas, USA)<sup>1</sup> equipped with an RTK-level AgGPS<sup>®</sup> 442 GNSS receiver (Trimble Navigation Ltd., Sunnyvale, California, USA). Then, maps of shallow (0–30 cm)  $EC_a$  and field elevation were used to determine nine field locations with the expectation that water demands would differ. Corresponding communication nodes for a wireless sensor network monitoring soil matric potential and temperature at four depths (18, 48, 79 and 109 cm) were installed in each of these nine locations.

### Sensor installation locations

To define soil water content monitoring sites,  $EC_a$  and the field elevation maps shown in Fig. 2 were analyzed using three combined optimization criteria (Adamchuk et al. 2011). These criteria included: (1) complete spatial field coverage using the S-optimality criterion; (2) even distributions throughout both data layers ( $EC_a$  and field elevation in this

<sup>1</sup> Mention of a trade name, proprietary product, or company name is for presentation clarity and does not imply endorsement by the authors, the University of Nebraska-Lincoln, or McGill University, nor does it imply exclusion of other products that may also be suitable.

case) using the D-optimality criterion; and (3) the relative homogeneity of the selected sites using a criterion based on the sum of squared differences between measurements obtained in each location and its immediate neighbors. The overall objective function was the geometric mean of these criteria normalized against the median of a large number of random selections (Adamchuk et al. 2010).

As a result, the same number of locations represented three physical parts of the field divided by two waterways (nodes 1, 4 and 7; nodes 2, 5 and 8; nodes 3, 6 and 9) and the same number of locations represented areas with low (nodes 2, 3 and 4) and high (nodes 5, 6 and 7) EC<sub>a</sub> as well as low (nodes 1, 2 and 3) and high (nodes 7, 8 and 9) field elevations (Fig. 3). The northeastern location (node 9) represented a non-irrigated area while the rest of nodes were distributed across irrigated land. However, it should be noted that the end gun covered the northwestern location (node 7), the southeastern location (node 3) was at the pivot stopping position and the remaining nodes represented different distances from the center of the irrigation pivot system (Fig. 4).

#### Sensor network installation

As shown in Fig. 5, each monitoring location was equipped with a wireless communication node (eN2100 eKo node, Crossbow Technology, Inc., San Jose, California, USA). Each sensor array was comprised of four thermistors and four soil matric potential sensors (in this case, Watermark<sup>®</sup> Granular Matrix sensors) with their centers placed 18, 48, 79 and 109-cm below the soil surface. Each sensor was intended to represent a 30-cm layer of soil. Soil matric potential is related to the energy that must be available in plants to extract water stored in the soil profile. In such a sensor, electrical resistance between the two electrodes imbedded in a sensor responds to water moved through a transmission material with a consistency close to that of fine sand. The sensors were calibrated at 21 °C. Measurements conducted using thermistors installed at the same depth were used to adjust soil matric potential measurements according to Thompson et al. (2006):

$$\psi_{adj} = \psi + 0.018(T - 21) \tag{1}$$

where  $\psi_{adj}$  is the temperature-corrected soil matric potential (kPa);  $\psi$  is the original soil matric potential measurement (kPa); and  $T$  is the soil temperature (°C). Both  $\psi$  and  $\psi_{adj}$  values are negative.

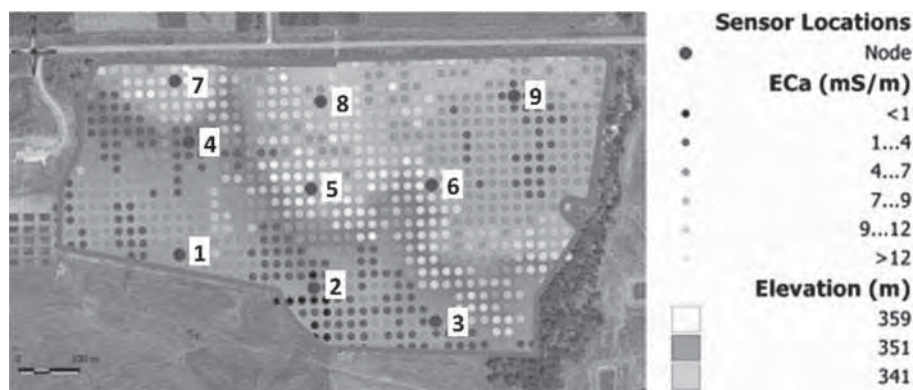


Fig. 2 Maps of topsoil EC<sub>a</sub> and field elevation with soil water monitoring sites

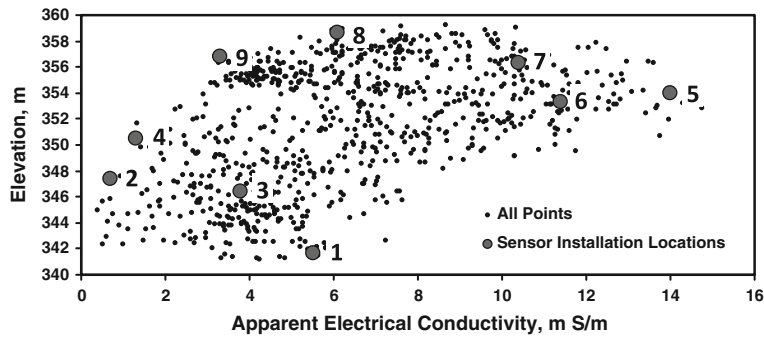


Fig. 3 Relationship between  $EC_a$  and field elevation

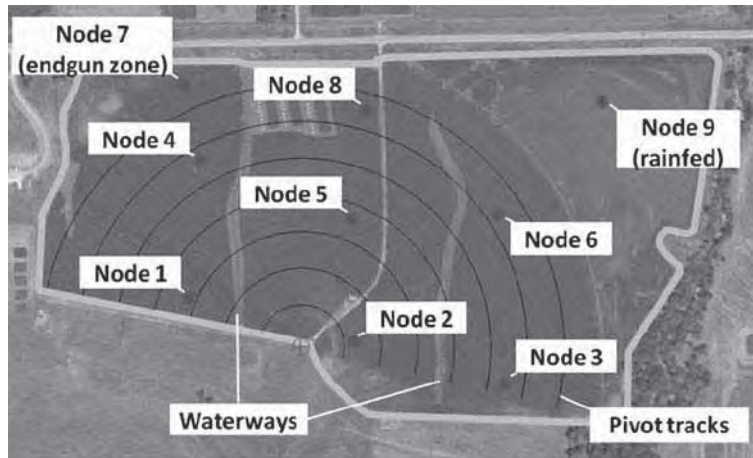
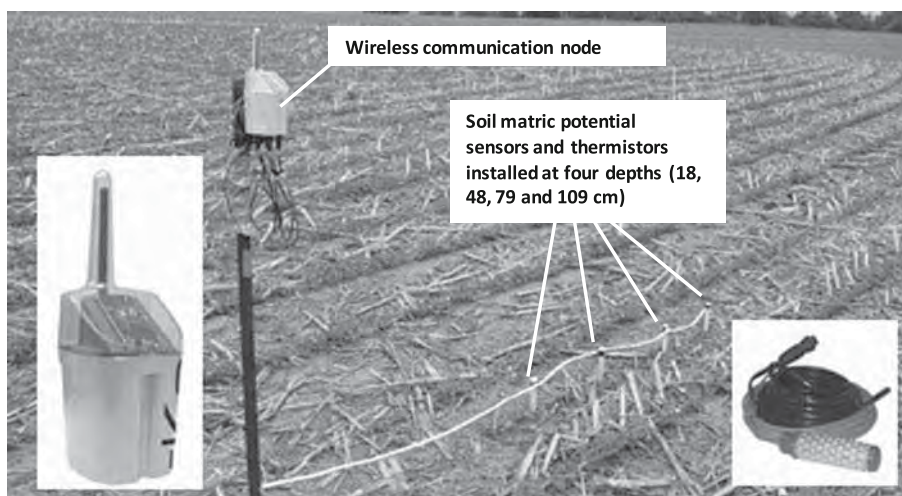


Fig. 4 Geometry of the center-pivot irrigation system with selected monitoring sites

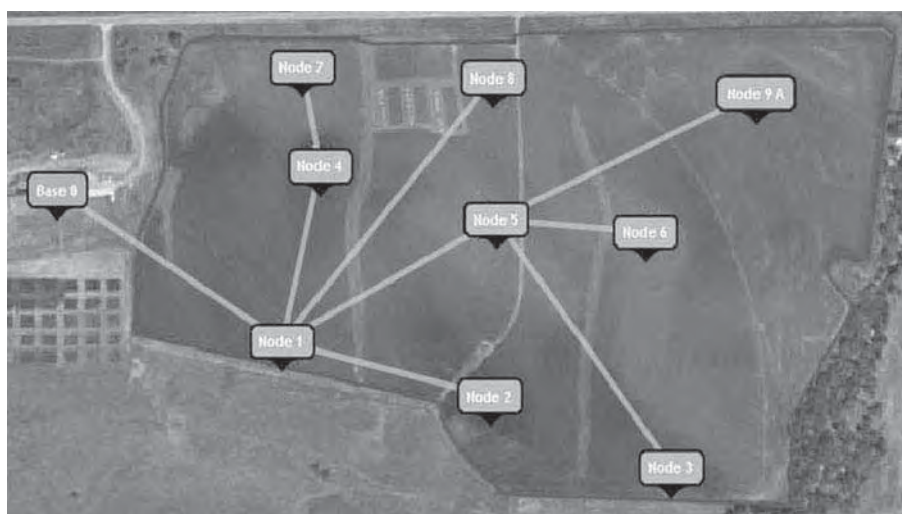
All the nodes were connected in a wireless sensor network (Fig. 6) with a 2.4 GHz communication frequency supporting the 802.15.4/ZigBee standard. All radios were bidirectional (half-duplex) and supported a range of up to 450 m using XMesh with low-power protocol. Each node worked as a repeater transmitting messages from other nodes that were within communication range. The eKo Gateway (eG2100, Crossbow Technology, Inc., San Jose, California, USA) was installed in a nearby building to store the data sent by each node in 15-min time intervals, as well as to view and manage data using the eKoView web interface. The eKo Base Radio (eB2110, Crossbow Technology, Inc., San Jose, California, USA) had an added hi-gain outdoor 15dBi omni-directional antenna (HAO15SIP, Hawking Technology, Inc., Irvine, California, USA) to ensure a reliable connection between the base and nodes. This extended the communication range to allow the base radio connect the nodes up to 3-km away.

In addition to monitoring the soil matric potential and temperature, node 10 was installed at the edge of the field to monitor atmospheric conditions (Fig. 7) and integrated into the network similarly to the other nodes. It was connected to an ambient temperature and humidity sensor (ES1201, Crossbow Technology, Inc., San Jose, California, USA) and





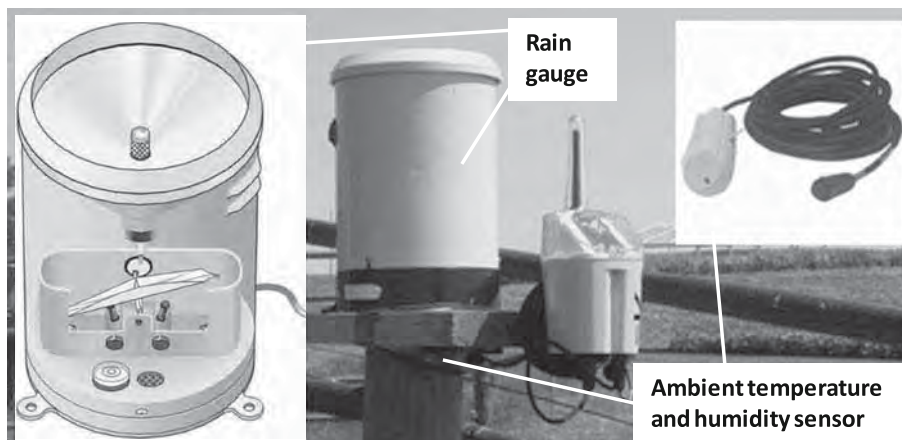
**Fig. 5** Soil water content monitoring site



**Fig. 6** The wireless sensor network

a tipping bucket rain gauge (TB3, Hydrological Services Pty. Ltd., Warwick Farm, New South Wales, Australia).

The irrigation schedule was followed as usual—according to soil conditions around node 8 as observed by an irrigation engineer. Based on personal communication, this location was used because of its relatively easy access and reasonable representation of the overall field conditions. A number of factors, including soil water content, were taken into account when scheduling irrigation. Experimental data were collected from 29 June to 4 October (both 2009 and 2010), which corresponds to the 14 weeks growing period with the highest water demand.



**Fig. 7** Atmospheric conditions monitoring station

#### Sensor network communication

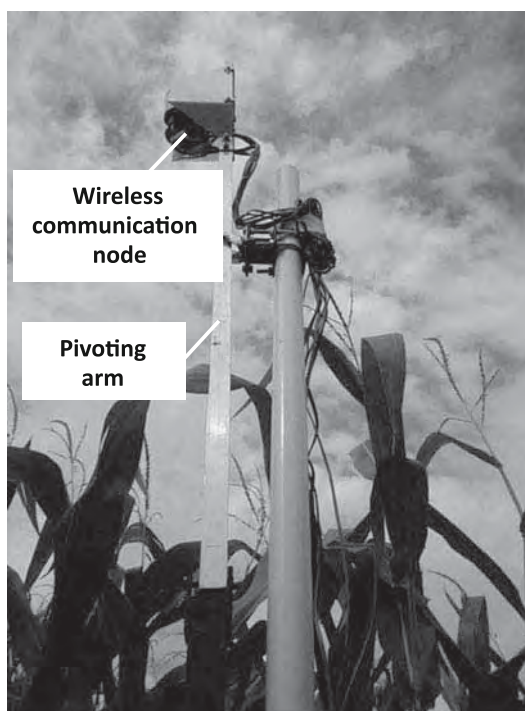
The experimental field was under corn-soybean rotation with soybean grown in 2009 and corn in 2010. The nodes were initially installed 1-m above ground, which was enough to ensure stable communication between nodes when soybean and young corn plants were growing. As the corn grew taller (beginning of July), the canopy affected data passage; a similar challenge has been described by Le Vine and Karam (1996) and Li et al. (2010). Therefore, pivoting arms (Fig. 8) were installed in 2010 to raise nodes 1 through 6 and node 8 to a height of 2.7 m to allow the irrigation system (with clearance of 2.5 m) operating without damaging the nodes. The other nodes (7, 9 and 10) were raised to a height of 3.4 m since they did not interfere with the irrigation system passage.

To quantify the effect of communication distance between a pair of nodes and the height of the corn canopy in our experimental field, a separate test was conducted. Three different transects as three replications were selected to monitor signal quality parameters at different distances from the node sending a signal. The influence of the height of the canopy was tested by moving the pair of communication nodes up and down in relation to the level of the top of the crop at five different positions (30-cm above crop canopy, level with the top of the crop, as well as 30, 60 and 90-cm below the top of the crop). In each trial, both receiving and transmitting nodes were placed close to each other and the receiving node was then moved away until communication was lost. The maximum communication distance, as well as the received signal strength indicator (RSSI) and link quality indicator (LQI), were recorded. For this study, as long as the nodes were able to communicate, the system was considered to be functional even with low RSSI and LQI. Thus, the maximum communication distance was the key parameter for future network designs.

#### Data analysis

All experimental data were collected and stored in comma-delimited text files. After all erroneous measurements (e.g. lost sensor communication or damaged sensor connections, which would show negative measurements) were removed, temperature compensation was applied according to Eq. 1.

**Fig. 8** A pivoting arm installed in the corn field in 2010



Although soil matric potential is an acceptable indicator of a crop's ability to extract soil water, it does not indicate how much water can be extracted. On the other hand, volumetric and gravimetric water content indicate the amount of water stored in the soil profile, but does not indicate accessibility of this water to plant roots. Therefore, our approach has been to combine both properties by monitoring soil matric potential normalized with respect to a site-specific threshold value that would be indicative of the potential for water stress. Although not well defined, many irrigation practitioners tend to look at the percentage of soil water depletion as an indicator of crop available water. Maintaining soil water depletion in a specified range can be viewed as an acceptable routine when it comes to irrigation water management (EPA 2003). It has been noted that 50 % depletion and more causes water stress in corn and maintaining the water content around the 20–30 % depletion range through the majority of the growing season is reasonable.

To implement this strategy, the Saxton model (Saxton and Rawls 2006) was used to define the relationships between soil matric potential and volumetric water content in each location and at a depth based on soil texture and organic matter content analyzed in a commercial lab:

$$\theta_V = (A/\Psi_\theta)^{\frac{1}{B}} \quad (2)$$

where  $\theta_V$  is the soil volumetric water content ( $\text{cm}^3/\text{cm}^3$ );  $A$  and  $B$  are soil-specific coefficients; and  $\Psi_\theta$  is the soil matric potential (kPa).

The coefficients  $A$  and  $B$  were found to fit a linear relationship between  $\ln(\Psi_\theta)$  and  $\ln(\theta_V)$  using estimated volumetric water content at  $\Psi = 33$  kPa and wilting point ( $\Psi = 1500$  kPa). Corresponding estimates of  $\theta_V$  were found based on percent clay, percent sand and organic matter measurements according to Saxton and Rawls (2006).



From the known value of volumetric water content, percent depletion was calculated as a fraction of volumetric water content corresponding to field capacity:

$$Depletion = \frac{\theta_{fc} - \theta_V}{\theta_{fc} - \theta_{wp}} \cdot 100\% \quad (3)$$

where *Depletion* is percent depletion;  $\theta_{fc}$  is the volumetric water content at field capacity ( $\text{cm}^3/\text{cm}^3$ ); and  $\theta_{wp}$  is the volumetric water content at wilting point ( $\Psi = 1500$  kPa,  $\text{cm}^3/\text{cm}^3$ ).

Based on the estimates of field capacity summarized by Schwab et al. (1993) for a variety of soil textures, the surface of field capacity for the entire soil texture triangle was interpolated with 1 % resolution and used to determine the volumetric water content and soil matric potential values at field capacity for each location using the measured soil particle size distribution. Volumetric water content values at wilting point ( $\Psi = 1500$  kPa) were estimated according to Saxton and Rawls (2006). These data are summarized in Table 1.

In addition, plants do not extract water uniformly throughout their rooting depth (Taylor and Klepper 1973). Root development is variable throughout the growing season. Based on the estimates of rooting depth and the plant root zone extraction pattern rule (Hoffman et al. 2007), the weight of measurements from the sensors installed at four depths through the soil profile was distributed according to the growth stage.

A water stress index (WSI) was estimated for each node by using the following equation:

$$WSI = \sum_{i=1}^4 w_i \cdot \left( \frac{\psi_i}{\psi'_i} - 1 \right) \quad (4)$$

where  $w_i$  is the weight of water extracted in the plant root zone for depth  $i$ ;  $\Psi_i$  is the soil matric potential measurement at  $i$ th depth (kPa); and  $\Psi'_i$  is the soil matric potential at certain threshold level of depletion (25 % depletion was used in this research as half of the most frequently cited percent depletion causing water stress in corn) at  $i$ th depth (kPa).

This index is an example of a unitless value that relates real-time soil matric potential and a site-specific threshold value that indicates the “normal” level of soil water depletion. Naturally,  $WSI < 0$  means water sufficient conditions and potential for water stress would cause  $WSI > 0$ . To develop real-time maps of the overall field water stress status, a simple linear regression method was used to predict WSI using field elevation and apparent soil electrical conductivity at each point in time:

$$WSI = \beta_0 + \beta_1 \cdot EC + \beta_2 \cdot Elev + \beta_3 \cdot EC \cdot Elev \quad (5)$$

where EC is measured apparent soil electrical conductivity (mS/m); *Elev* is field elevation (m);  $\beta_0$  through  $\beta_3$  are model coefficients.

Although measurements were collected with a 15-min time interval, daily averages of WSI were used to demonstrate this approach, which resulted in 99 independent regression models with a new set of coefficients each day. Since non-irrigated parts of the field (node 9) had a different rate of water supply from the irrigated parts (the rest of monitoring sites), only the first eight nodes were used to establish the linear regression model each day.

Finally, the percentage of field area under different levels of water stress (wet, no stress, low stress, high stress) was determined each day. This information served as the main

**Table 1** Soil texture, volumetric water content ( $\theta_v$ ) and soil matric potential ( $\Psi$ ) at each node location and depth

Node	Depth (cm)	Clay (%)	Sand (%)	Soil type	Wilting point	Field capacity	
					$\theta_v$ (cm <sup>3</sup> /cm <sup>3</sup> )	$\theta_v$ (cm <sup>3</sup> /cm <sup>3</sup> )	$\Psi$ (kPa)
1	18	20	41	Loam	0.14	0.33	13.0
1	48	26	34	Loam	0.17	0.36	14.0
1	79	24	39	Loam	0.16	0.33	14.7
1	109	21	45	Loam	0.13	0.30	14.2
2	18	4	90	Sand	0.02	0.09	13.5
2	48	6	88	Sand	0.04	0.10	15.0
2	79	6	91	Sand	0.03	0.08	14.0
2	109	4	91	Sand	0.02	0.08	14.8
3	18	18	57	Sandy loam	0.12	0.24	25.7
3	48	14	65	Sandy loam	0.10	0.21	16.2
3	79	10	79	Sandy loam	0.06	0.14	17.6
3	109	6	90	Sand	0.03	0.09	16.1
4	18	7	80	Loamy sand	0.05	0.12	19.9
4	48	6	85	Loamy sand	0.04	0.11	17.3
4	79	5	86	Loamy sand	0.03	0.09	15.7
4	109	2	91	Sand	0.01	0.07	15.8
5	18	34	15	Silty clay Loam	0.21	0.36	41.8
5	48	30	14	Silty clay Loam	0.19	0.34	48.8
5	79	23	17	Silty loam	0.15	0.31	38.7
5	109	15	15	Silty loam	0.10	0.27	48.7
6	18	36	17	Silty clay loam	0.22	0.37	37.4
6	48	39	14	Silty clay loam	0.24	0.39	34.5
6	79	42	10	Silty clay	0.25	0.38	46.6
6	109	29	22	Clay loam	0.18	0.33	38.2
7	18	33	16	Silty clay loam	0.21	0.36	42.7
7	48	34	16	Silty clay loam	0.21	0.36	39.6
7	79	30	15	Silty clay loam	0.19	0.34	45.7
7	109	22	13	Silty loam	0.14	0.31	44.5
8	18	27	20	Silty loam	0.18	0.34	36.3
8	48	35	16	Silty clay loam	0.21	0.36	38.3
8	79	43	14	Silty clay	0.26	0.39	41.3
8	109	33	14	Silty clay loam	0.20	0.35	41.9
9	18	25	16	Silty loam	0.16	0.33	42.3
9	48	33	14	Silty clay loam	0.20	0.35	42.9
9	79	34	12	Silty clay loam	0.21	0.36	43.1
9	109	36	14	Silty clay loam	0.22	0.37	37.9

output that could be used to optimize irrigation scheduling. Although analysis was completed during both growing seasons, only 2009 data have been illustrated in this publication as 2010 data had a number of gaps caused by network communication.

A generalized linear regression was applied to define the water stress value at every field location and at every point in time. To achieve this, Eq. 5 was extended to include the average WSI value for all eight locations:

$$\begin{aligned}
 WSI = & \beta_0 + \beta_1 \cdot EC + \beta_2 \cdot Elev + \beta_3 \cdot \overline{WSI} + \beta_4 \cdot EC \cdot Elev \\
 & + \beta_5 \cdot EC \cdot \overline{WSI} + \beta_6 \cdot Elev \cdot \overline{WSI} + \beta_7 \cdot EC \cdot Elev \cdot \overline{WSI}
 \end{aligned}
 \tag{6}$$

where  $\overline{WSI}$  is the daily average values of WSI of all eight nodes (node 1–8);  $\beta_0$  through  $\beta_7$  are model coefficients.

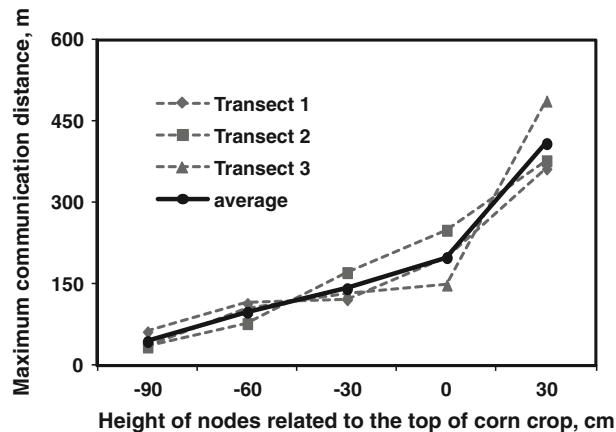
### Results and discussion

From the quality communication test, it was shown that the estimated maximum communication distances were significantly influenced by the relative height of the nodes with respect to the top of the corn canopy for all three transects (Fig. 9). It appeared that communication distances greater than 150 cm were feasible only when the nodes were less than 30-cm below the top of the crop. An unobstructed line of sight between the nodes allowed for up to a 450-m range of communication. After constructing pivoting arms described earlier (Fig. 8), the network was able to function reliably.

Figures 10 and 11 illustrate soil matric potential (logarithmic scale) and volumetric water content corresponding to the field capacity and a 25 % depletion for each depth at all nine locations. Based on the Saxton and Rawls (2006) model,  $\ln(\Psi_\theta)$  and  $\ln(\theta_v)$  have a linear relationship. As shown in Fig. 12, within the range of  $\Psi =$  field capacity to 300 kPa, the relationship between  $\ln(\Psi_\theta)$  and  $\theta_v$  can be assumed linear for all depths at each location other than node 4 and node 2 which had an extremely low value of volumetric water content.

Figures 13 and 14 show the measurements of soil matric potential at four depths in each of the nine node locations with cumulative water supply information for 2009 and 2010, respectively. Based on the results in 2009, the measurements of soil matric potential were relatively low at the beginning of the growing season (wet soil) with gradual drying until the middle of August when a substantial rainfall event occurred, after which the “drying

**Fig. 9** Maximum communication distance in response to the relative height of the nodes with respect to the top of the crop



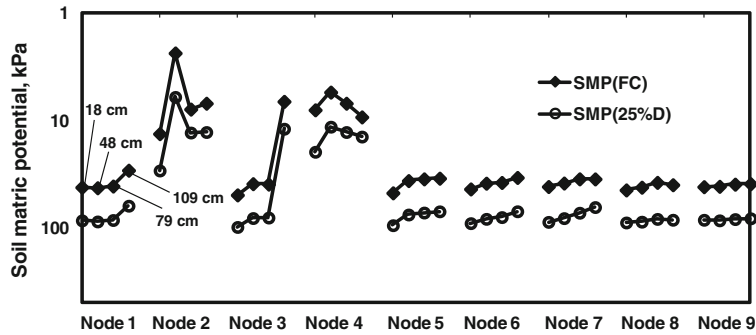


Fig. 10 Soil matric potential values at field capacity and 25 % soil water depletion

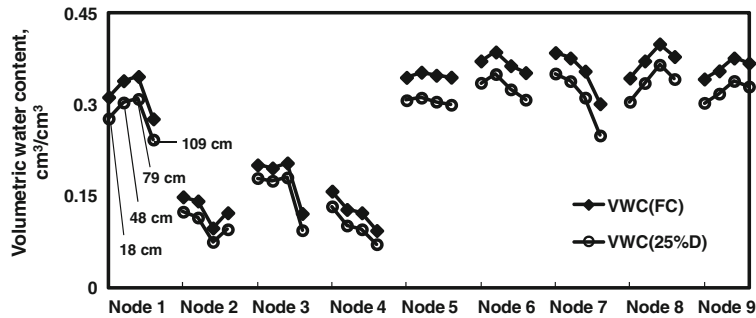


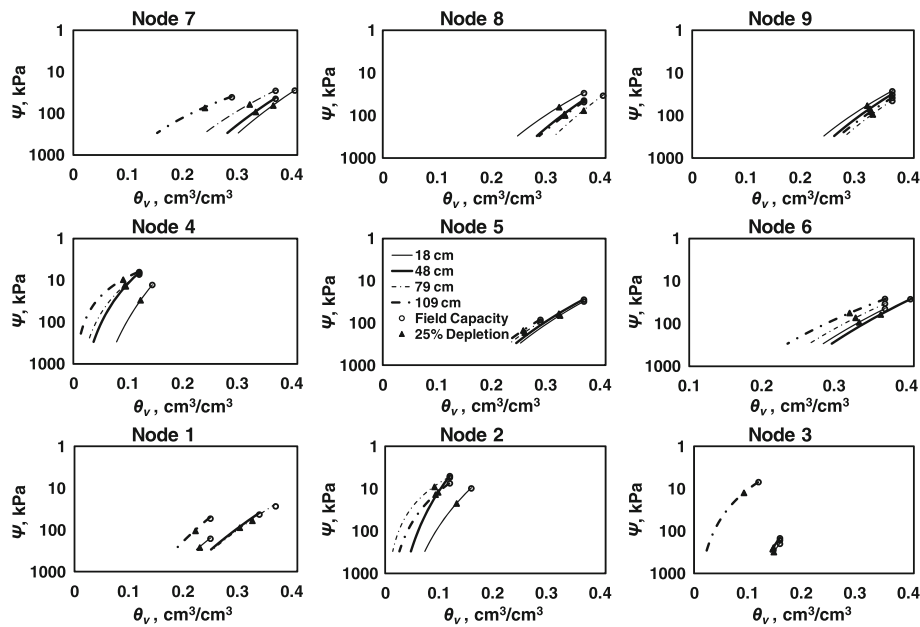
Fig. 11 Volumetric water content values at field capacity and 25 % soil water depletion

out” process started again. Also, as one would expect, deeper soil layers dried slower than shallower soil layers. Based on the expected range of soil matric potential and the measurements obtained, most of the field did not experience severe water stress (percent depletion above 50 %). Node 3 was located in sandy soil, which contributed to relatively high values of soil matric potential near the soil surface. During the sensor installation at node 3, a compacted sand pan was discovered at approximately 70-cm below the soil surface, which may have restricted root development, caused more water extraction from shallow depths and contributed to wetter soil deeper in the profile.

Also, it is important to note that at node 9 (a rainfed location with fine-textured soil) installed sensors did not exhibit a significant shortage of water during the entire growing season. Apparently, August rainfalls were sufficient to recharge the soil profile for the remainder of the growing season, which explains why irrigation was terminated early in the season. It appears that had precipitation been lower, severe depletion could have been reached near the end of August.

The trend in soil water conditions was not as obvious for 2010 as a result of numerous losses of data, as shown in Fig. 14. However, it was observed that measurements collected at node 3 (especially the second depth) showed a trend of water stress towards the end of the data collection. In contrast to 2009, there was no major precipitation from late July to mid-August. Sensors installed in irrigated areas other than node 3 (compacted sand pan at 70-cm deep) did not indicate water stress through the entire season. Node 9 (non-irrigated area) exhibited an increasing shortage of water; however, it was not strong enough to cause significant yield losses.

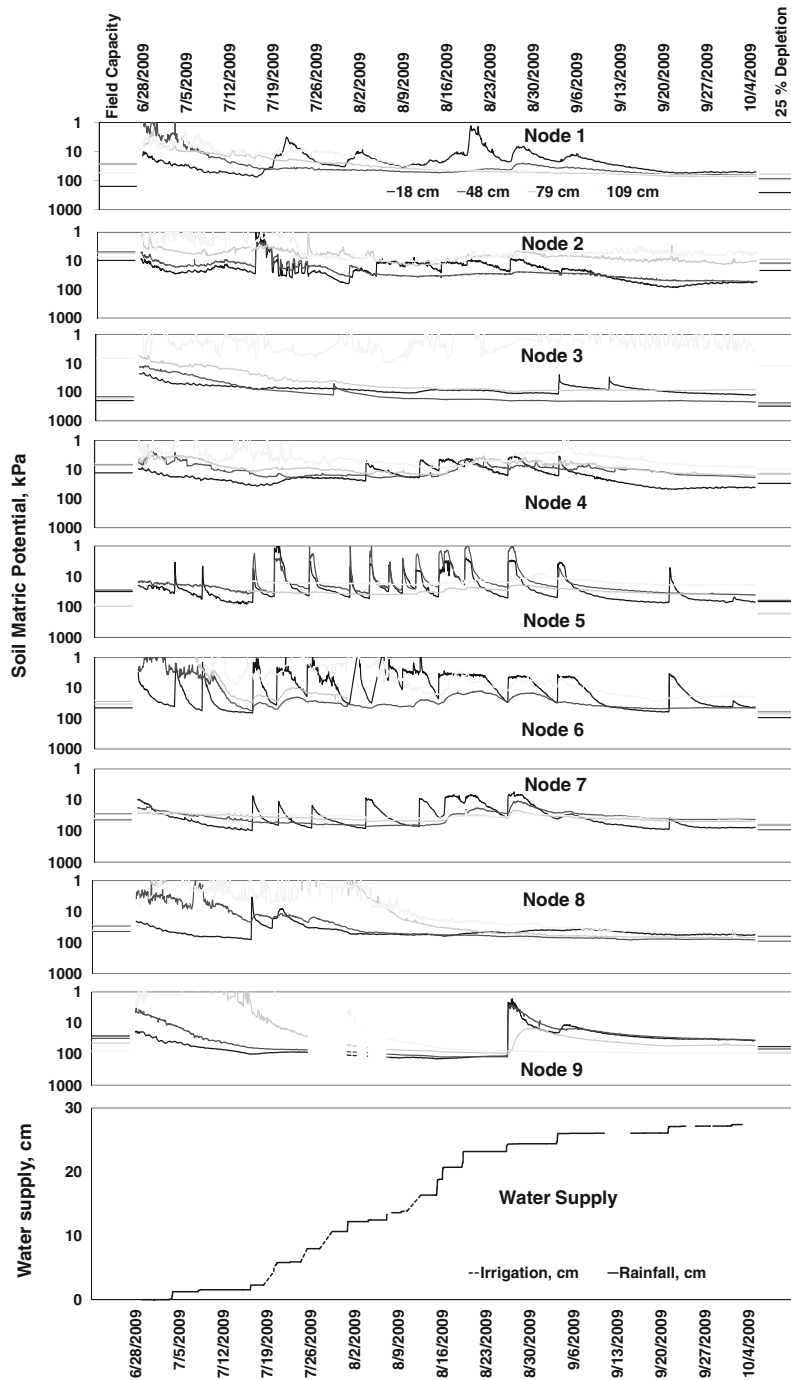




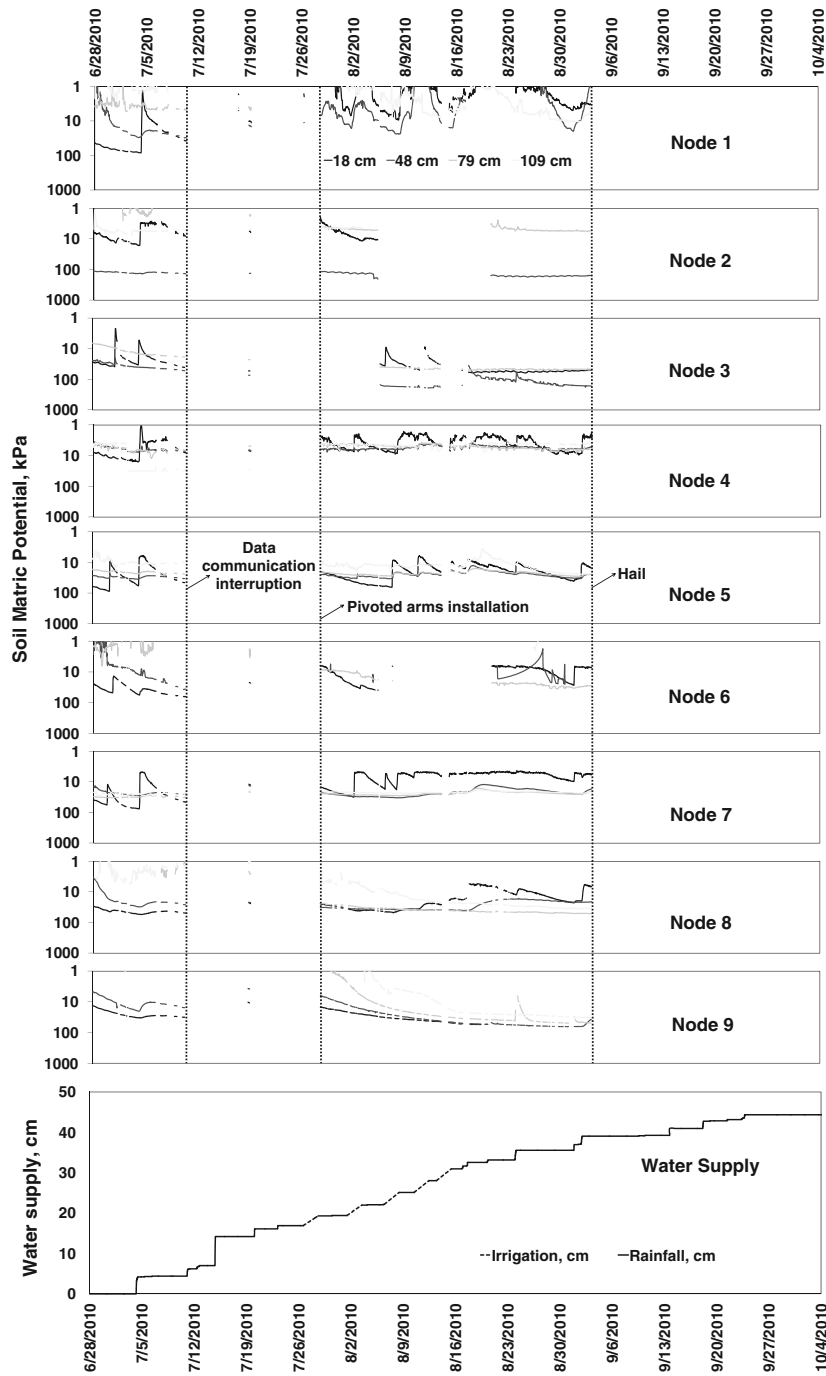
**Fig. 12** The relationship between soil matric potential  $\Psi$  and volumetric water content  $\theta_v$  ( $\Psi$  ranges from field capacity to 300 kPa)

Based on Hoffman et al. (2007), Fig. 15 illustrates the assumed change in soybean root depth through the growing season. Calculated according to Eq. 4, the daily average values of *WSI* were modeled for each day using Eq. 5. Fig. 16 illustrates the coefficients of linear regression and corresponding  $R^2$  values are shown in Fig. 17. All the coefficients of linear regression, including  $\beta_3$  (product of EC and Elev) were found to be significant  $\alpha = 0.05$ ). Apparently, the highest  $R^2$  value of the regression models was 0.97 (Fig. 18) corresponding to the beginning of August and the lowest  $R^2$  was 0.63 at the end of August. Considering this field, node 2 (with average *WSI* value of 0.14, which was much higher compared to other sites) was the most influential node. Once again, this emphasized the importance of diversified water storage conditions being represented by a monitoring site. For less variable fields, the coefficient of determination of linear regression models may not be as high as for this site.

Figure 19 illustrates the *WSI* maps for the entire field on three different days. The maps clearly show the difference in soil water availability. To summarize such maps obtained for each of the 99 days, Fig. 20 shows the change of *WSI* distributions over time. In this case “wet soil” was assumed to have *WSI* values less than  $-1$ ; “no stress” with *WSI* value ranging between  $-1$  and  $0$ ; “low stress” with *WSI* value between  $0$  and  $1$  (over 25 % depletion); “high stress” with *WSI* value larger than  $1$  (soil matric potential is twice greater than 25 % depletion). It appears that at the beginning of the growing season, the entire field was under wet conditions and the percentage of wet soil continued to decrease until the first irrigation was applied in July. These distributions can provide key input when it comes to irrigation scheduling and deciding on site-specific water management. For example, an irrigation engineer can base decisions on the irrigation schedule by combining weather forecast and percentage of field area going into “low stress” and “high stress” categories. Site-specific water management should be considered when simultaneously, a significant

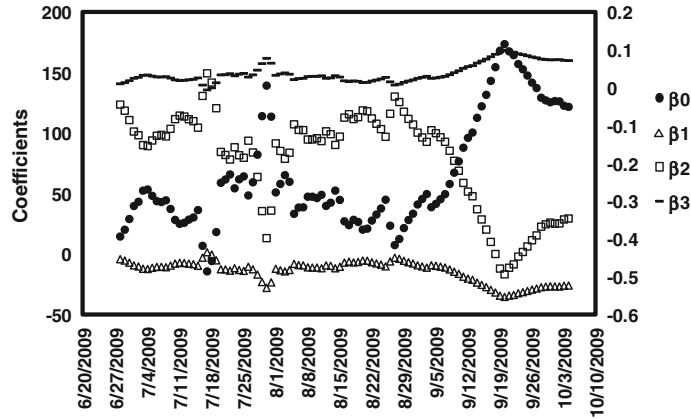
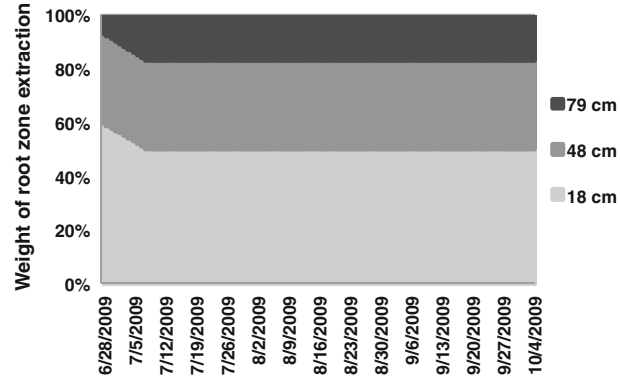


**Fig. 13** Seasonal records of soil matric potential with respect to field capacity and 25 % soil water depletion and water supply (2009)



**Fig. 14** Seasonal records of soil matric potential with respect to field capacity and 25 % soil water depletion and water supply (2010)

**Fig. 15** Weight of root water extraction through soil profile for soybean



**Fig. 16** The coefficients of WSI regression model ( $\beta_0$  and  $\beta_1$  were plotted on primary axis while  $\beta_2$  and  $\beta_3$  were plotted on secondary axis) during the growing season

portion of land indicates negative WSI values while another significant part of the field experiences water stress.

As shown in Fig. 16, the coefficient of simple linear regression (Eq. 5) changes over time according to a pattern induced by water supply and evapotranspiration rates. Since field average WSI is directly affected by this pattern, the generic regression model (Eq. 6) resulted in the following equation with all the coefficients significant ( $\alpha = 0.05$ ) and  $R^2 = 0.82$ :

$$\begin{aligned}
 WSI = & 122.2 - 25.4 \cdot EC - 0.35 \cdot Elev + 164.8 \cdot \overline{WSI} + 0.072 \cdot EC \cdot Elev \\
 & - 32.5 \cdot EC \cdot \overline{WSI} - 0.47 \cdot Elev \cdot \overline{WSI} + 0.092 \cdot EC \cdot Elev \cdot \overline{WSI} \quad (7)
 \end{aligned}$$

Once more, this proves consistency in relative water availability. In other words, relatively dry parts of the field will most likely remain dryer than relatively wet parts of the field with water supply and demand continuously scaling this difference. Once such a relationship is established, the overall decision-making process will rely on quantitative assessments of potential water stress impact in terms of field area subjected to different levels of water availability. The ability to derive a well-determined regression model for the entire growing season suggests the capability to apply the concept using a much smaller



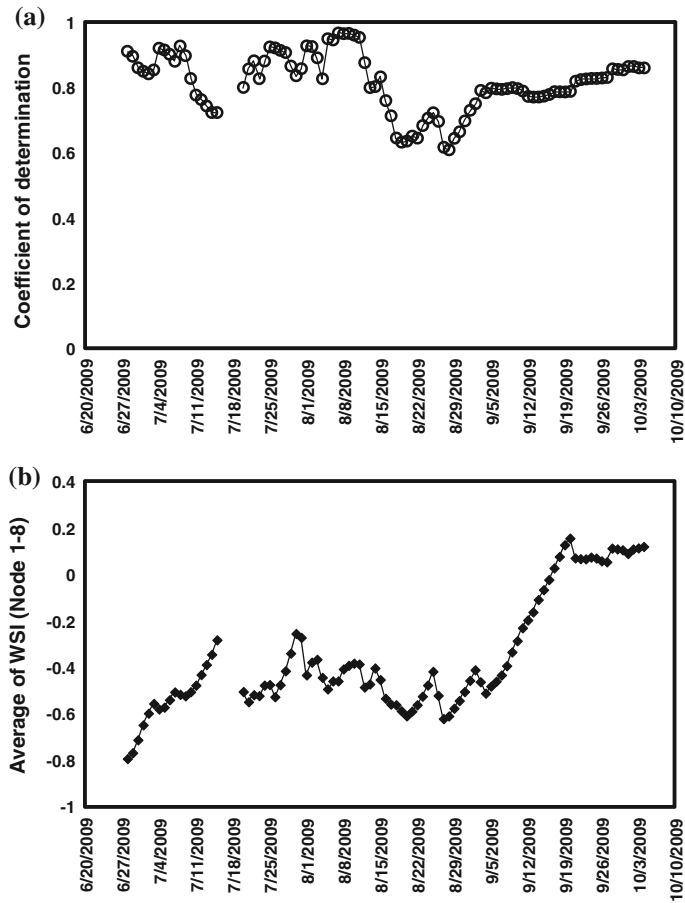
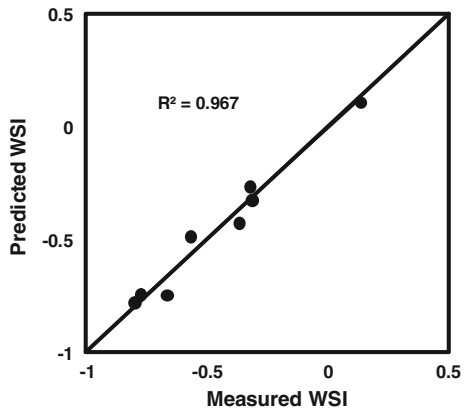
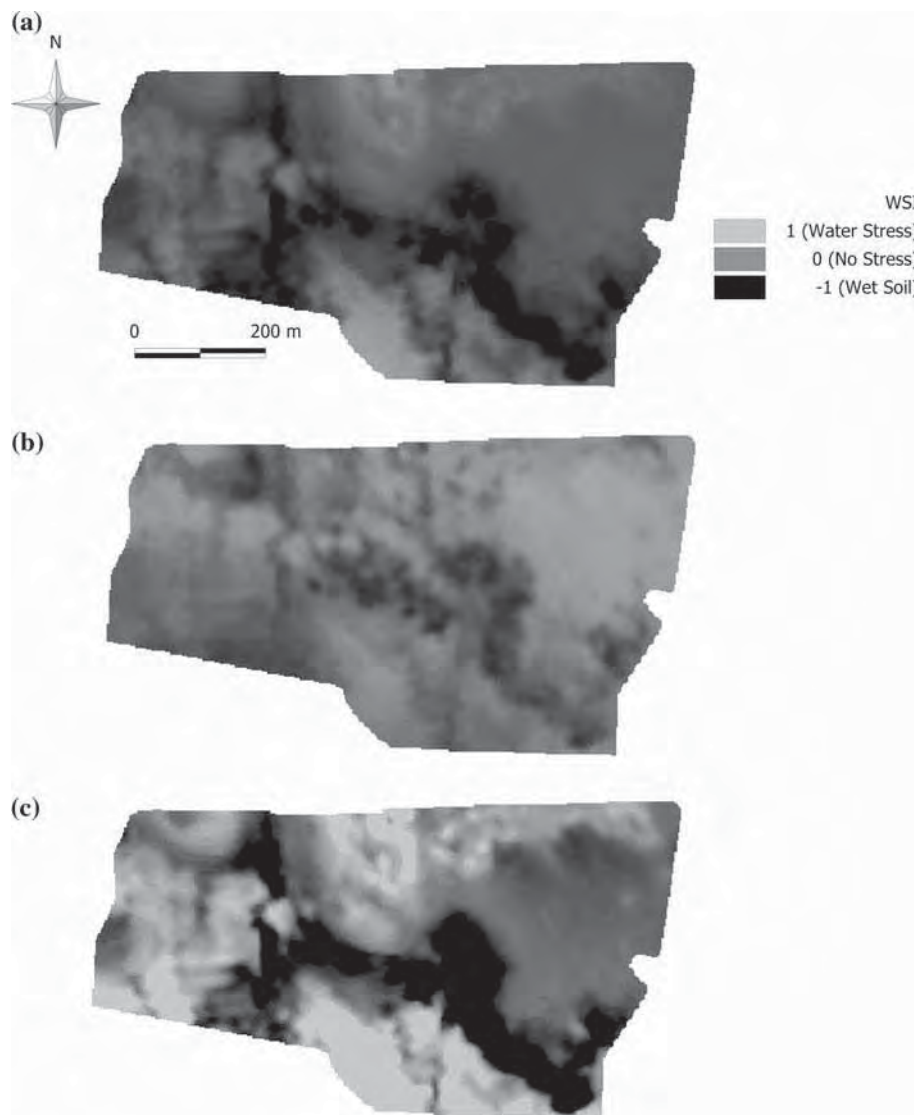


Fig. 17 Coefficients of determination (a) and field average of WSI (b) during the 2009 growing season

Fig. 18 Example of a simple linear regression used to predict WSI (08/10/2009)





**Fig. 19** Examples of WSI maps produced on 7 Jul (a), 14 Aug (b) and 20 Sep (c)

number of nodes. As long as the monitoring sites represent both, relatively wet and dry parts of the field, real time prediction of water stress is feasible. Our follow-up work has focused on the assessment of WSI prediction quality in response to the number and placement of soil water content monitoring sites.

### Conclusions

In this research,  $EC_a$  and field elevation maps were used to locate a wireless network of nine nodes to monitor soil matric potential and temperature at four depths for soybean

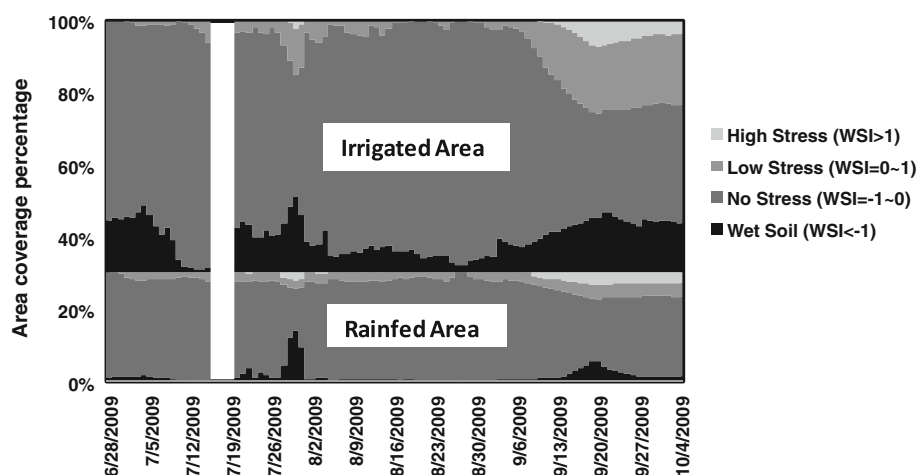


Fig. 20 Areal distribution of WSI through the 2009 growing season

grown in a 37-ha production field with a center pivot irrigation system. The network provided the capability for visualizing and storing data remotely. A soil water retention model, combined with measured estimates of soil texture and organic matter content, was used along with soil matric potential and soil temperature measurements to predict the volumetric water content and, ultimately, the depletion of available water throughout the growing season. Due to sufficient precipitation, significant water stress was not present and the need for VRI management was not obvious. However, it was clear that coarse-textured soils located predominantly along the lower field elevations had water regimes that differed from the rest of the field. This approach allowed for day-to-day monitoring of the area of the field most susceptible to water stress to assess its need for irrigation. Geographical confounding of relatively dry and wet field areas may suggest the need for a VRI practice.

**Acknowledgments** This publication is a contribution of the University of Nebraska Agricultural Research Division, supported in part by funds provided through the Hatch Act and by funds provided through the Nebraska Center for Energy Science Research through the Water, Energy and Agriculture Initiative (WEAI).

## References

- Adamchuk, V. I., Hummel, J. W., Morgan, M. T., & Upadhyaya, S. K. (2004). On-the-go soil sensors for precision agriculture. *Computers and Electronics in Agriculture*, 44(1), 71–91.
- Adamchuk, V. I., Pan, L., Marx, D. B., & Martin, D. L. (2010). Locating soil monitoring sites using spatial analysis of multilayer data. In *Proceedings of 19th World Congress of Soil Science*. Brisbane, Australia, 1–6 August 2010. International Union of Soil Science.
- Adamchuk, V. I., Viscarra Rossel, R. A., Marx, D. B., & Samal, A. K. (2011). Using targeted sampling to process multivariate soil sensing data. *Geoderma*, 163(1–2), 63–73.
- Bao, Y., Wu, Y., He, Y., & Wang, L. (2004). *Study on signal processing for a crop irrigation virtual instrument system*. Paper No. 701P1004. St. Joseph, MI: ASABE.
- EPA. (2003). Irrigation water management. In *National Management Measures to Control Nonpoint Source Pollution from Agriculture*. EPA 841-B-03-004 Technical Guidance (pp. 4-157–4-202). Washington, DC: Environmental Protection Agency.
- Han, Y. J., Khalilian, A., Owino, T. O., Farahani, H. J., & Moore, S. (2009). Development of Clemson variable-rate lateral irrigation system. *Computers and Electronics in Agriculture*, 68(1), 108–113.

- Han, W., & Wu, P. (2011). Derivation and application of hydraulic equation for variable-rate contour-controlled sprinklers. *African Journal of Biotechnology*, 10(79), 18214–18221.
- Hedley, C., & Yule, I. (2009). A method for spatial prediction of daily soil water status for precise irrigation scheduling. *Agricultural Water Management*, 96(12), 1737–1745.
- Hedley, C., Yule, I., Tuohy, M., & Vogeler, I. (2009). *Key performance indicators for variable rate irrigation implementation on variable soils. Paper No. 096372*. St. Joseph, MI: ASABE.
- Hoffman, G. J., Evans, R. G., Jensen, M. E., Martin, D. L., & Elliott, R. L. (2007). *Design and operation of farm irrigation systems* (2nd ed.). St. Joseph, MI: ASABE.
- Kim, Y., & Evans, R. G. (2009). Software design for wireless sensor-based site-specific irrigation. *Computers and Electronics in Agriculture*, 66(2), 159–165.
- Kim, Y., Evans, R. G., & Iversen, W. M. (2009). Evaluation of closed-loop site-specific irrigation with wireless sensor network. *Journal of Irrigation and Drainage Engineering*, 135(1), 25–31.
- King, B. A., & Kincaid, D. C. (2004). A variable flow rate sprinkler for site-specific irrigation management. *Applied Engineering in Agriculture*, 20(6), 765–770.
- Le Vine, D. M., & Karam, M. A. (1996). Dependence of attenuation in a vegetation canopy on frequency and plant water content. *IEEE Transactions on Geoscience and Remote Sensing*, 34(5), 1090–1096.
- Li, Z., Hong, T., Wang, N., & Wen, T. (2010). Data transmission performance for 2.4 GHz in-field wireless sensor network. *2nd International Conference on Computer Engineering and Technology*, 1, 465–469.
- Miranda, F. R., Yoder, R., & Wilkerson, J. B. (2003). A site-specific irrigation control system. Paper No. 031129. St. Joseph, MI: ASABE.
- Omary, M., Camp, C. R., & Sadler, E. J. (1997). Center pivot irrigation system modification to provide variable water application depths. *Applied Engineering in Agriculture*, 13(2), 235–239.
- Saxton, K. E., & Rawls, W. J. (2006). Soil water characteristic estimates by texture and organic matter for hydrologic solutions. *Soil Science Society of America Journal*, 70(5), 1569–1578.
- Schwab, G. O., Fangmeier, D. D., Elliot, W. J., & Frevert, R. K. (1993). *Soil and water conservation engineering* (4th ed.). New York: Wiley.
- Taylor, H. M., & Klepper, B. (1973). Rooting density and water extraction patterns for corn. *Agronomy Journal*, 65(6), 965–968.
- Thompson, R. B., Gallardo, M., Agüera, T., Valdez, L. C., & Fernández, M. D. (2006). Evaluation of the Watermark sensor for use with drip irrigated vegetable crops. *Irrigation Science*, 24(3), 185–202.
- Vellidis, G., Tucker, M., Perry, C., Kvien, C., & Bednarz, C. (2008). A real-time wireless smart sensor array for scheduling irrigation. *Computers and Electronics in Agriculture*, 61(1), 44–50.



# VERTICAL SOIL PROFILING USING A GALVANIC CONTACT RESISTIVITY SCANNING APPROACH

*Luan Pan, Viacheslav I. Adamchuk and Shiv Prasher, McGill University, Montreal, QC, Canada*

*Robin Gebbers, Leibniz-Institute for Agricultural Engineering (ATB), Potsdam, Germany*

*Richard S. Taylor, DUALEM, Milton, ON, Canada*

## Abstract

Site-specific crop productivity is heavily influenced by the ability of soil to store water and nutrients while still permitting excess water to drain. One of the most promising practices to define spatial soil heterogeneity in terms of physical characteristics is the application of electrical conductivity/resistivity maps. The instruments used to map such soil characteristics use galvanic contact and capacitive coupling resistivity measurements as well as electromagnetic induction. Despite the type of instrument, the geometrical configuration between signal transmitting and receiving elements defines the shape of the depth response function. To assess vertical variation of soils from the surface, many modern instruments use multiple transmitter-receiver pairs to record electrical conductivity/resistivity signals applicable to different soil depths. Alternatively, vertical sounding methods can be used to measure a change in apparent soil electrical conductivity with the depth at a specific location. This paper examines the possibility for dynamic assessment of soil profiles using a surface galvanic contact resistivity scanning approach, with transmitting and receiving electrodes configured in an equatorial dipole-dipole array. An automated scanner system has been developed and tested in the agricultural field environment with different soil profiles. While operating in the field, the distance between current injecting and measuring pairs of rolling electrodes was varied continuously from 40 to 190 cm. The resulting scans were evaluated against 1-m deep soil profiles and that of an electromagnetic induction instrument at various depths.

**Keywords:** galvanic contact resistivity, apparent soil electrical conductivity, on-the-go sensing

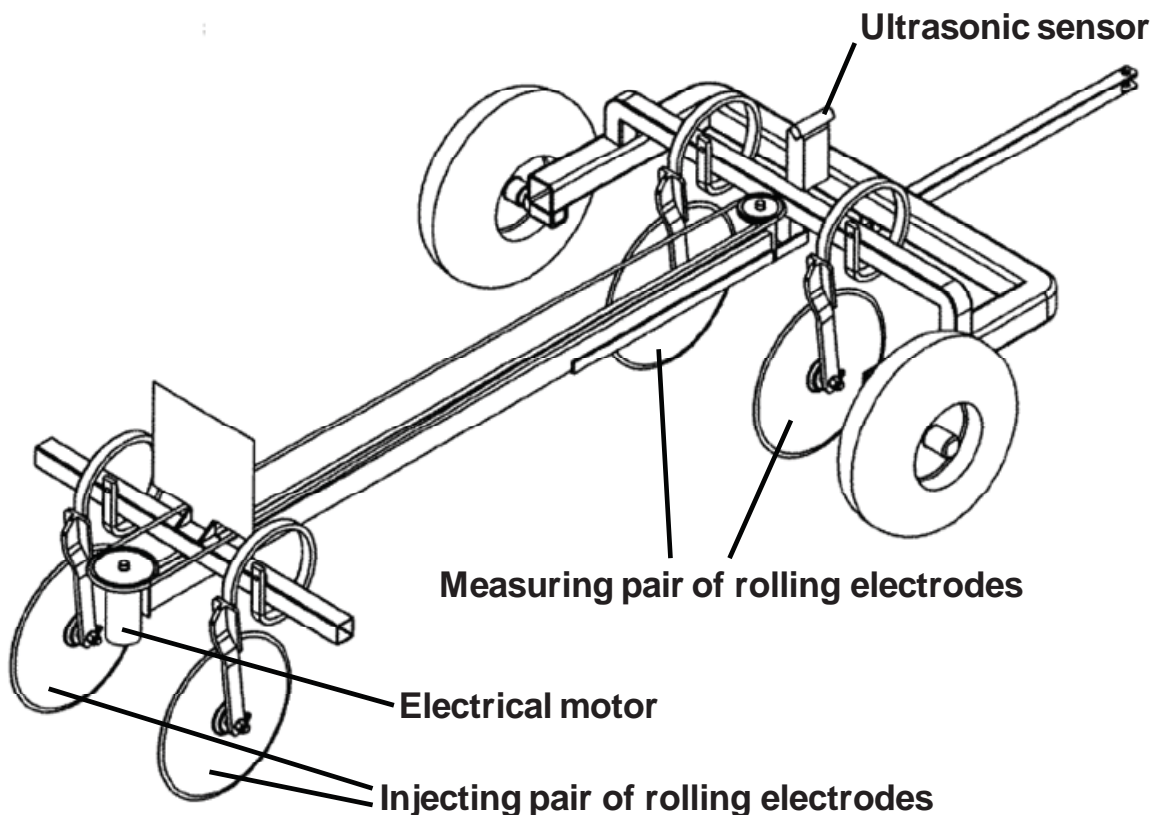
## Introduction

The ability of soil to conduct, or accumulate, electrical charge has been linked with several physical and chemical soil properties. The reliability and ease of operation made soil electrical conductivity/resistivity mapping instruments popular when used to describe lateral soil heterogeneity. However, a number of applications require understanding of how soil profiles change with depth. For example, the consistency and depth of topsoil can be related to the ability of the soil profile to store water and plant nutrients. Knowing this information can help optimize nutrient and water use efficiency through differentiated approaches to applying mineral fertilizers and irrigation water according to local needs. Traditional point-based explorations are laborious and are not economical when it comes to high-resolution soil mapping. Proximal soil sensing technology used to map soil electrical conductivity/resistivity has been accomplished through galvanic contact and capacitive coupling resistivity techniques (Allred et al., 2008) as well as electromagnetic induction (Daniels et al., 2008). Several different methods have been used to estimate how soil electrical characteristics change with depth. For instance, based on the summarized practices with existing equipment (Gebbers et al., 2009), Lueck & Ruehlmann (2012) developed a sensor system which allowed the measurements of electrical

parameters at four depth levels to a maximum of 1.5 m, using six pairs of rolling electrodes. The objective of this study was to develop a prototype galvanic contact resistivity system capable of obtaining continuous scans of soil electrical resistance, using two pairs of rolling electrodes configured in equatorial dipole-dipole mode with variable distances between the two pairs.

## Materials and methods

To obtain the apparent electrical conductivity/resistivity values of soil, artificially generated electric current was supplied to the soil using an injecting pair of electrodes; the resulting potential difference was obtained using a measuring pair of rolling electrodes (Figure 1). Veris Q series disc electrodes (Veris Technologies, Inc., Salina, Kansas, USA) were used to build the system. It was equipped with an electrical motor displacing the movable pair of electrodes in an automated mode. A 4-point light hp (Lippmann Geophysikalische Messgeräte, Schaufling, Germany) instrument, with the frequency setting of 25 Hz, was used to conduct the measurements. The distance between the current injecting and measuring pairs of rolling electrodes was varied continuously from 40 to 190 cm. The distance within each pair was fixed at 45 cm. An ultrasonic sensor was used to measure the distance between the two pairs of electrodes while a Global Positioning System (GPS) receiver measured the geographic coordinates. All the data were recorded using a developed LabView (National Instruments, Inc., Austin, Texas, USA) interface. With the current speed of operation, it took less than 10 s for the complete displacement of the movable pair of electrodes in both directions while operating either in a stationary mode or on-the-go.



**Figure 1:** Prototype galvanic contact resistivity scanner

With equatorial dipole-dipole configuration, the soil electrical resistivity and conductivity can be calculated using the following equations (Roy & Apparao, 1971):

$$\rho = \frac{\pi}{\sqrt{L} \sqrt{(0.45^2 + L^2)}} \cdot \frac{U}{I} \quad (1)$$

$$\sigma = \frac{1000}{\rho} \quad (2)$$

where  $\rho$  is electrical resistivity ( $\Omega \cdot m$ );  $L$  is the distance between the injecting and the measuring pairs, m;  $U$  is the measured voltage (mV);  $I$  is the generated current (mA);  $\sigma$  is electrical conductivity (mS/m)

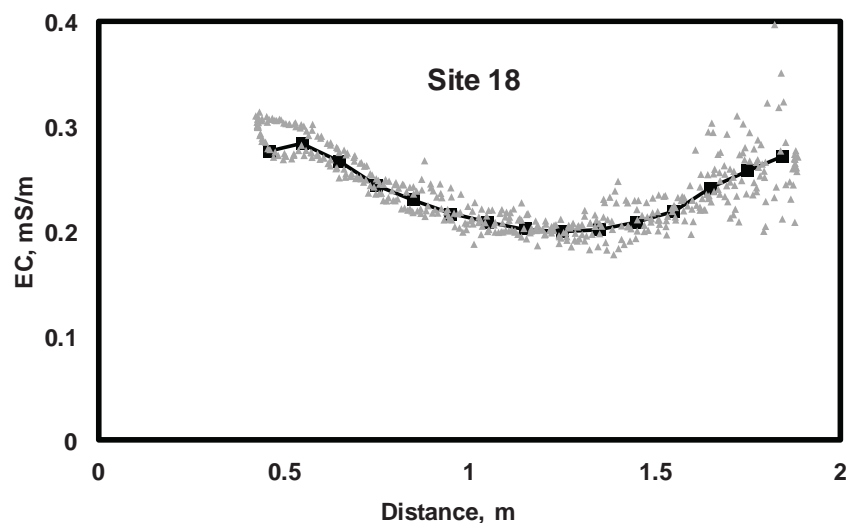
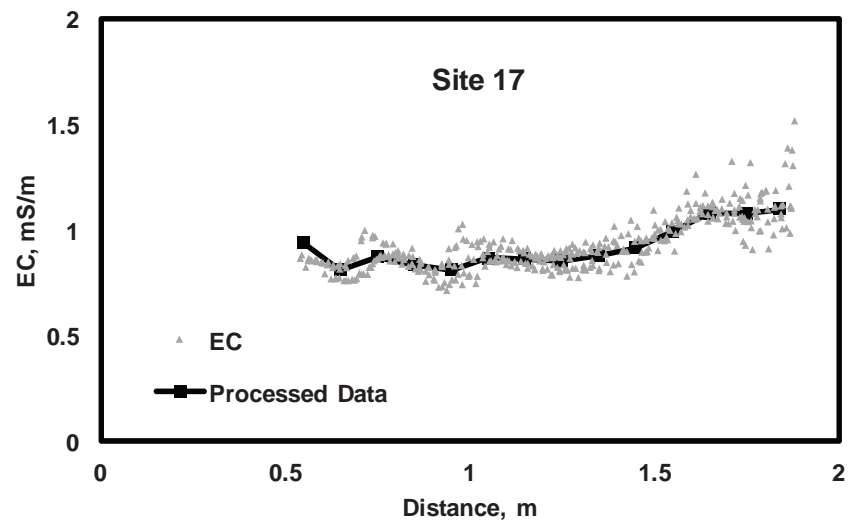
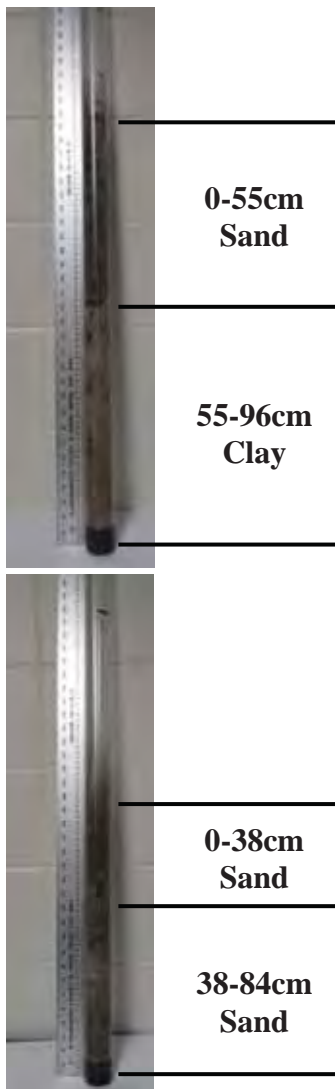
This scanner system has been tested in three agricultural fields at Macdonald farm of McGill University. In each location, the measurements of the scanner system were compared with those of the DUALEM-21S (DUALEM, Milton, ON, Canada) sensor using a multilayer perception (MLP) neural network algorithm with IBM-SPSS Modeler Version 15.0 (IBM-SPSS, Inc., Chicago, Illinois, USA). DUALEM measurements were interpreted according to a dual-layer soil profile model (depth of the top layer as well as the electrical conductivity of both layers of soil). The processed data from the developed scanner system (fifteen weighted average measurements from 40 to 190 cm distance between electrode pairs calculated at 10-cm increment) were assigned as inputs, while three parameters for a DUALEM sensor profile were set as targets to be predicted. The 41 total scans (from 23 locations with replicates in most of them) were used for both training and testing, given the setting of 2 hidden layers with 5 elements each. Ten trials were run following the same procedure. Simple linear regressions were applied to correlate the predicted parameters and those estimated using the DUALEM sensor. In addition, 1-m deep soil cores were obtained in the same locations. The soil cores were divided into subsamples based on the soil color. Soil texture was measured for each section using a hydrometer method. In this paper, four soil profiles were illustrated along with the corresponding scans.

## Results and discussion

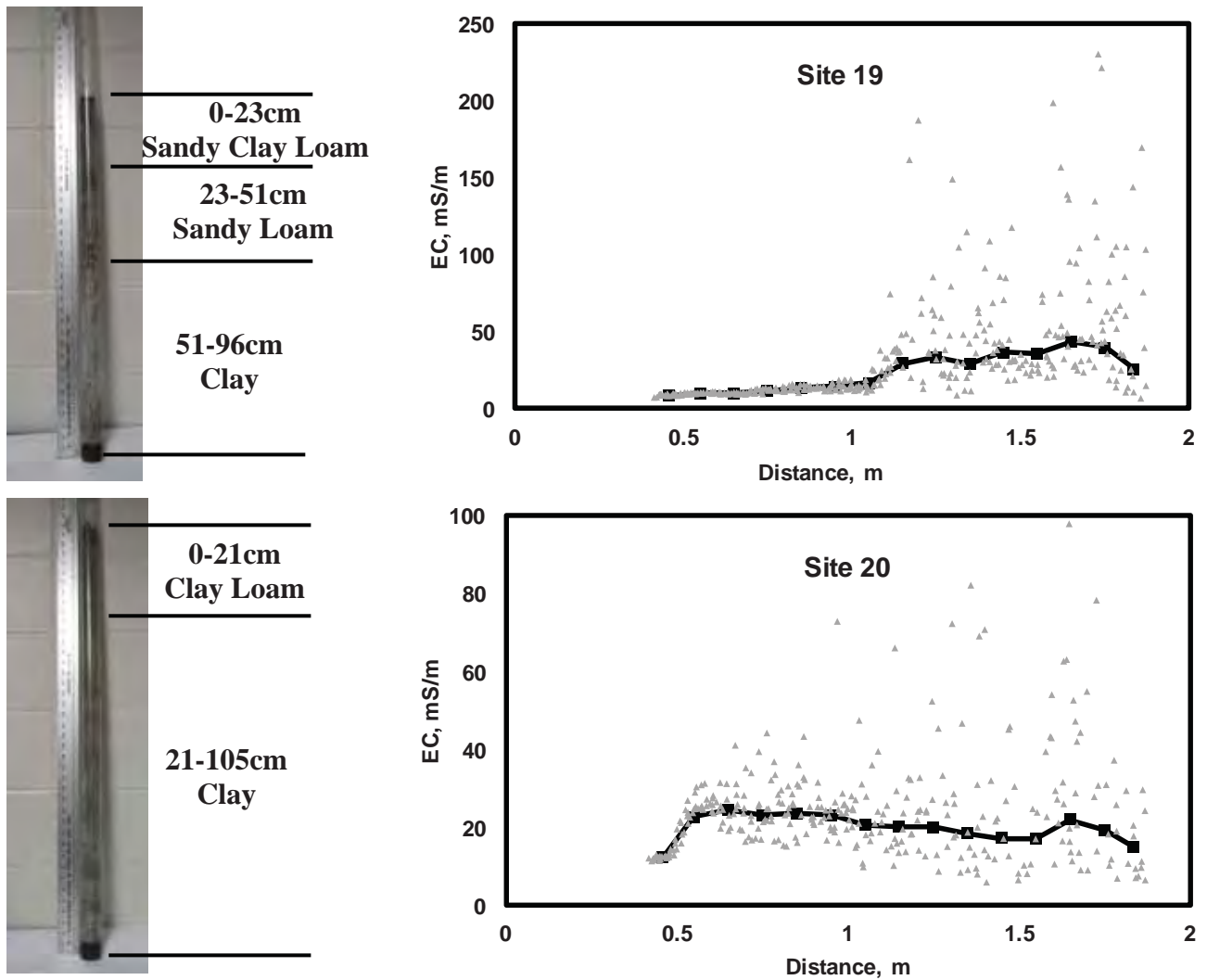
While analyzing the scanning system output, it was observed that the data had a relatively high noise caused by both electrical and mechanical inconsistencies. The magnitude of this noise increased at larger distances between injecting and measuring pairs of electrodes. Therefore, data processing involved filtering out outliers and weighted data averaging. Then, both equations 1 and 2 were applied to estimate electrical resistivity and conductivity. Figure 2 illustrates four soil profiles along with the corresponding electrical conductivity scans. It appeared that the shapes of these scans were consistent with the observed change in soil texture with depth. Thus, sandy site 18 produced a relatively constant scan with low  $EC_a$  values. Sites 19 and 20 indicated higher  $EC_a$  values with greater distances between the injecting and measuring pairs, which was consistent with an increase in clay content with depth. Unfortunately, deeper soil profiles would be needed to validate the quality of the  $EC_a$  scans inversion, which is the next step in developing a scanning system data processing procedure.

Figure 3 illustrates corresponding soil EC profiles with parameters inverse-modeled from DUALEM-21S data. Since the DUALEM-21S has four depths of investigation, a three-parameter (two-layer) soil-profiling model was used. Based on the results of the ten trials using an MLP neural network algorithm, significant correlations between scanner predictions and DUALEM estimates were observed

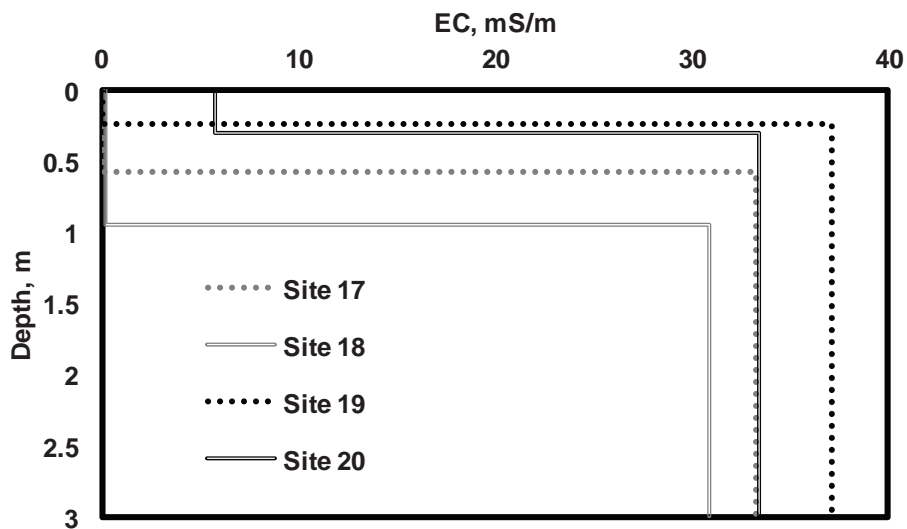
(Figure 4). However, this test did not illustrate robustness of predictions as a greater number of different soil profiles would be necessary to enable a proper validation procedure. In general, the scanner data corresponded to the estimations of the depth and conductivities of two-layer models using corresponding DUALEM-21S data. Furthermore, with more complex soil profile models, it appears that the developed scanning procedure could provide the greater number of  $EC_a$  measurements within the depth-range of interest. However, actual measurements of soil EC deeper than 1 m were not feasible at this time. In any case, it is clear that continuous scanning of galvanic contact soil resistivity could be applied to predict a change of soil physical consistency with depth, which indirectly could indicate the local potential of the soil profile to store water and nutrients or lose them through deep percolation and leaching. Further work is needed to evaluate the practical applicability of the platform developed and to optimize the operational and data interpretation parameters.



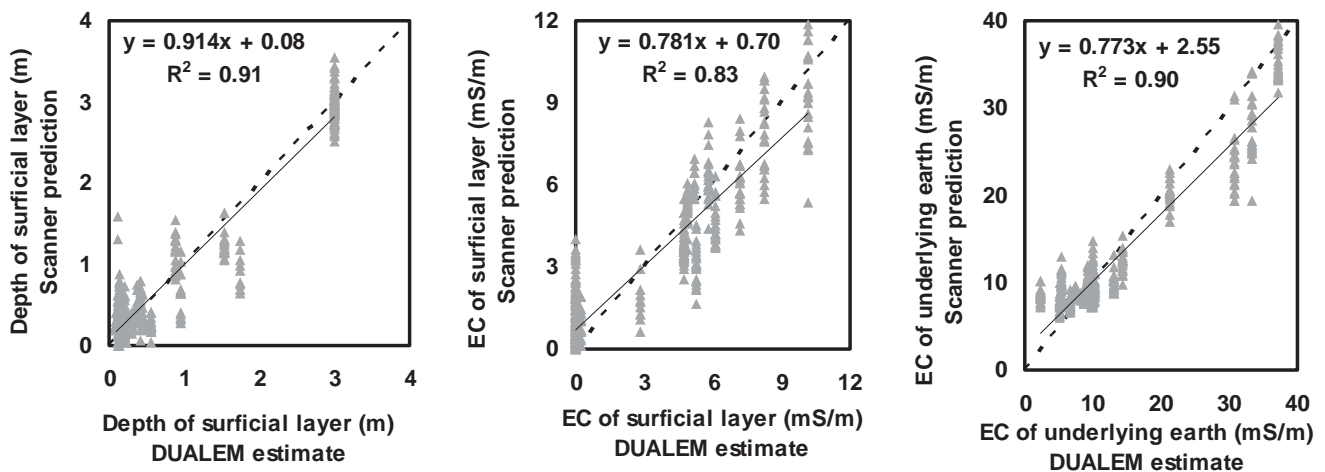




**Figure 2:** Four example soil profiles (left side) along with corresponding electrical conductivity scans (right side)



**Figure 3:** Soil profiles with inversion results of the DUALEM measurements for the four example sites



**Figure 4:** Relationships between scanner predictions and DUALEM estimates when treating the entire dataset as calibration data

## Conclusions

An automated scanner system with an equatorial dipole-dipole array has been developed and tested in several agricultural fields. Examples of 1-m soil cores indicated changes in soil texture with depth consistent with the scans obtained. However, deeper cores are needed to develop a proper methodology to interpret scanner measurements. On the other hand, obtained scans could predict two-layer model parameters (depth of surficial soil and EC of the two soil layers) estimated using a popular DUALEM-21S instrument. The advantage of the scanner developed is the quasi-continuous variation of depth sensitivity and lower cost as compared to the multi-electrode galvanic coupled sensors with fixed geometries. This approach is similar to vertical sounding measurements obtained by changing the height of an electromagnetic induction sensor above the ground.

## References

- Allred, B.J., Groom, D., Ehsani, M.R., Daniels, J.J., 2008, Resistivity methods. In: Allred, B.J. et al. (eds.) *Handbook of Agricultural Geophysics*. Boca Raton, Florida: CRC Press, pp. 85-108.
- Daniels, J.J., Vendl, M., Ehsani, M.R., Allred, B.J., 2008, Electromagnetic induction methods. In: Allred, B.J. et al. (eds.) *Handbook of Agricultural Geophysics*. Boca Raton, Florida: CRC Press, pp. 109-128.
- Gebbers, R., Lück, E., Dabas, M., Domsch, H., 2009, Comparison of instruments for geoelectrical soil mapping at the field scale, *Near Surface Geophysics*, 7, 179-190.
- Lueck, E., Ruehlmann, J., 2012, Resistivity mapping with GEOPHILUS ELECTRICUS - Information about lateral and vertical soil heterogeneity. Geoderma (in press).  
<http://dx.doi.org/10.1016/j.geoderma.2012.11.009>
- Roy, A., and Apparao, A., 1971, Depth of investigation in direct current methods, *Geophysics*, 36, 943-959.

# Real time sensing and ISOBUS in agriculture

**J. Lenz**, John Deere, Technology Innovation Center, Champaign, Illinois, USA;

**P. Münch**, John Deere, European Technology Innovation Center, Kaiserslautern, Germany;

**V. Adamchuk**, Bioresource Engineering Department, McGill University, Ste-Anne-de-Bellevue, Quebec, Canada

## Abstract

This paper describes a new option for varying agricultural inputs according to local needs by using an embedded, tractor-based controller. This function allows entry-level producers and machinery operators to implement variable rate technology based on their intuitive knowledge of their fields. Its versatility and open concept will allow tractor operators to define their rules in a way similar to setting up an A-B line typically used for navigation. In this case, a tractor controls its working parameters (e.g., travel speed, hydraulic flow, PTO engage, etc.) in response to selected internal, or external, sensors according to the parameters set up by the operator in real-time. The map-based, adds-on function will also allow the use of traditional georeferenced prescription maps.

## Introduction

One of the main objectives of smart farming, also known as precision agriculture, is to optimize crop production and reduce the environmental footprint by using new digital technologies [1]. In comparison to the wide-spread use of global navigation satellite system (GNSS) based guidance of field machinery, the adoption of variable rate technology has been somewhat limited and has yielded mixed results [2]. Lack of robust decision support mechanisms has been a major limitation when it comes to varying fertilization and seeding rates according to local needs [3]. Furthermore, the capital costs involved in this technology prevented a number of small and medium farmers from adopting its use. However, many producers understand the need for non-uniform field treatments. Thus, a farmer can identify distinct field areas where uniform application rates would not be optimal due to the limited, or excessive, crop productivity potential. It is our goal to equip these farmers with a technology that would match tractor operation parameters with local conditions according to the rules established by the operator.

## Materials and methods

Various components of a modern tractor intercommunicate using a controlled area network (CAN) binary unitary system (BUS). To improve the interconnectivity among different brands of farm machinery, the communication protocol has been standardized under the patronage of the International Organization for Standardization (ISO) [4]. This ISOBUS technology offers a number of still unrealized opportunities [5]. A tractor CAN has access to different sensors installed on a tractor as well as having the capability to send special commands to internal controllers and actuators. Furthermore, different data conversion systems can be used to transform signals from external sensors and to create the output required by external controllers. This makes a typical tractor controller capable of interacting with internal and external components in real time. Figure 1 illustrates the main components of a typical ISOBUS implementation. Today, much of the emphasis has been placed on the interaction between the tractor and the electronic control units (ECUs) as well as on the in-cab virtual terminal (VT). In this project, it has been assumed that medium-size farming involves a number of implements that do not have an ECU. In this case, a simple virtual instrument application (VI app) could be added to a VT to obtain sensor measurements and, based on a user-defined transformation, control the tractor operation directly via the ISOBUS.

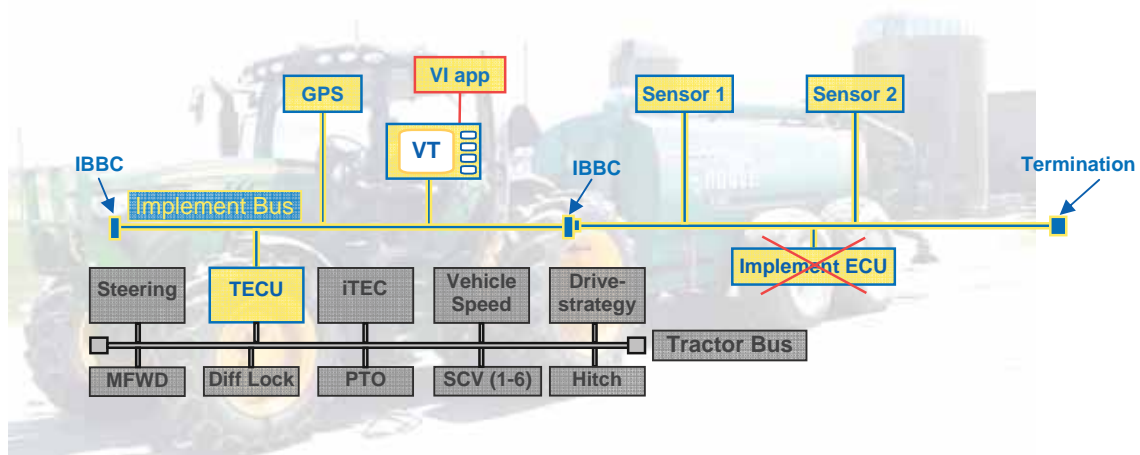


Fig. 1: Proposed ISOBUS tractor-based variable-rate control approach.

To illustrate the concept of the real-time tractor-based variable rate control strategy, let us assume that the operation parameter  $S$  (e.g. travel speed) is related to a status parameter  $E$  (field elevation). In our first example, this corresponds to the technique of reducing the amount of organic fertilizers applied at the lower field elevations to account for the potential runoff and to reduce the probability of contaminating nearby waterways.



Figure 2 illustrates two potential  $S=f(E)$  algorithms. In the first case (stage control option on the left), two critical values ( $E_a$  and  $E_b$ ) specify three intervals for  $E$  values that correspond to three different preset operation parameters  $S$ . Although this method is easy to implement, many control options (including travel speed) can be varied continuously in response to a changing input variable. The second case (proportional control option on the right) is an example of such an algorithm. In this case, any two points on the linear part of the functional relationship between  $E$  and  $S$  can be used to define the entire relationship.

In the case of the stage controller, default  $S_a$ ,  $S_b$ ,  $S_c$ ,  $E_a$  and  $E_b$  values can be easily changed during field operations. However, this is not as straightforward as with the proportional controller. Figure 3 illustrates a possible solution when a constant  $S$  is maintained until the six critical parameters ( $S_{min}$ ,  $S_a$ ,  $S_b$ ,  $S_{max}$ ,  $E_a$  and  $E_b$ ) are assigned. Both pairs of parameters,  $E$  and  $S$ , can be entered manually or taken from the measurements at a given location. Thus, when the tractor operator enters a field at a given field location requiring a relatively low or high  $S$  value, the system will record both parameters ( $E$  and  $S$ ) as soon as the operator clicks on either the A or the B button. When conditions change, the operator similarly clicks the other button. This is comparable to setting up an A-B line for automatic guidance. The settings should be completed as soon as possible to enable proper field operation through the rest of the field.

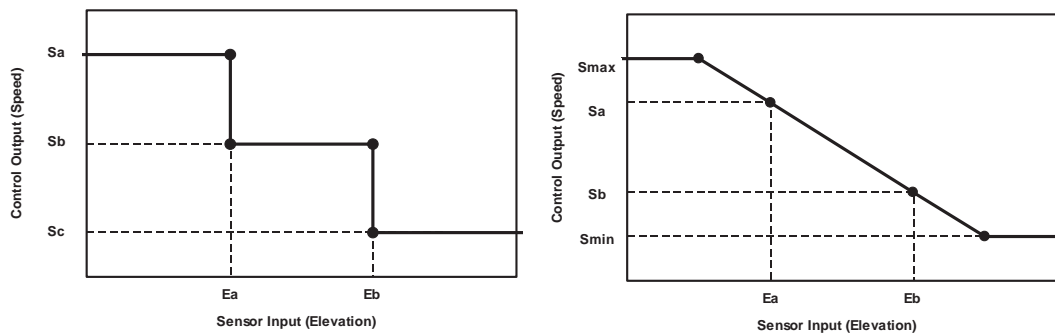


Fig. 2: Three stage control (left) and proportional (right) control algorithm.

To illustrate the first (three stage) control options, a John Deere 6430 tractor was equipped with a Sunnit PC with the newly developed application software that controlled travel speed using CANBUS (Figure 4). A StarFire 3000 GPS/GLONASS receiver can be used to determine the altitude and to feed this into the controller using ISOBUS communication. The setup

was used to apply liquid cattle manure in an 11-ha alfalfa production field at the Macdonald Farm of McGill University (Ste-Anne-de-Bellevue, Quebec, Canada).

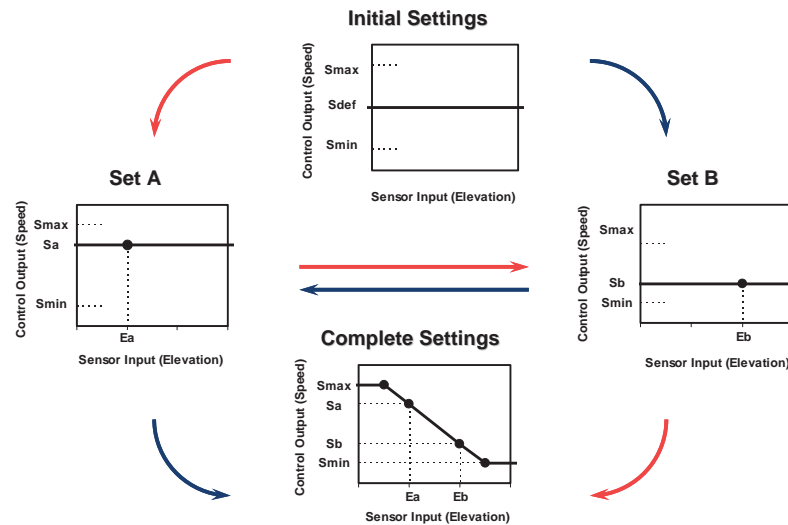


Fig. 3: Proportional controller set-up procedure.



Fig. 4: Experimental tractor (left), CANBUS connection (center) and a base station used to provide RTK-level GNSS signal to determine field elevation.

## Results and discussion

Figure 5 shows the record of travel speed calculated using a designated GPS receiver installed for monitoring purposes. Since wheel slippage did not change significantly across the field, the estimated travel speed appeared to range within 0.1 m/s with respect to the set value for the entire field. Figure 6 shows maps displaying log data recorded during the test. Reducing travel speed at higher elevations (areas with lower runoff expectations) could result in

higher fertilization rates. In this case, an increase of 18% of liquid cattle manure was disposed of in this field.

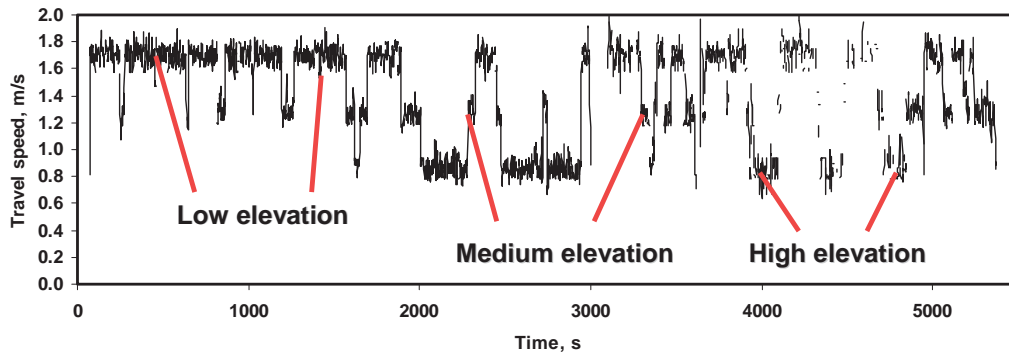


Fig. 5: External travel speed measurements during the test.

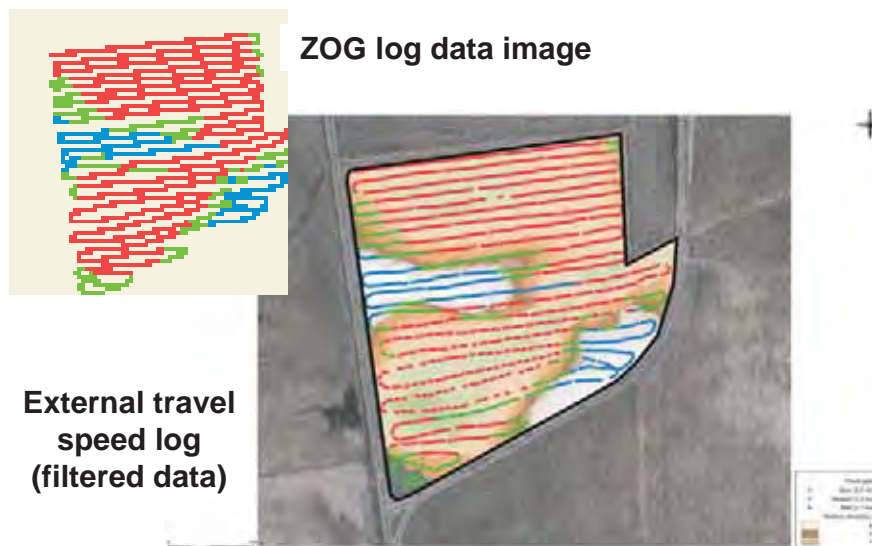


Fig. 6: Field test when the tractor was operated in automated speed control mode.

In addition to the distribution of liquid manure, our ongoing research is aimed at demonstrating the capabilities of tractor-based variable rate control when applying herbicides and harvesting forage crops. In the first case, travel speed is reduced when a relatively high level of undesired vegetation is detected optically in front of the tractor. While cutting the forage crop, travel speed is automatically reduced to maintain a steady flow of material in highly productive areas of the field, as assessed using an ultrasonic proximity sensor. Furthermore, we have implemented variable depth planting to account for spatially variable soil water content during planting. Thus, seeds are placed closer to the surface in wet conditions to allow for

uniform emergence. Certainly, many other traditional and innovative variable rate application scenarios can be implemented using the technology described in this study. Our adds-on option to use two internal and/or external sensor inputs and a predefined shape file make this technology versatile and suitable to build on existing site-specific crop management practices.

## **Conclusions**

At the current stage of precision agriculture, medium-size farm operators frequently resist adopting variable rate technology due to the cost of the equipment and the extra time required to learn the new technologies. On the other hand, many operators intuitively understand the need for differentiating their operations within a field in accordance with local needs. For decades, some operators have implemented what can be classified as VRT using manual control. What can be called a “smart tractor” concept will enable farmers to implement their intuitive practices in a more replicable and ergonomic way. The tractor will recognize the operator-defined conditions using internal or external sensors and it will replicate the operation settings across the field when conditions are similar. A combination of real-time sensor-based and map-based operations will add versatility. In general, the concept is similar to the principles of auto-guidance and will remove the gap between the use of traditional machinery and farming machinery equipped with high-end controllers and actuators.

## **References**

- [1] Gebbers, R. and V.I. Adamchuk. 2010. Precision agriculture and food security. *Science* 327(5967): 828-831.
- [2] Whipker, L.D. and J.T. Akridge. 2009. *Precision Agricultural Services. Dealership Survey Results*. Agricultural Center for Food and Business, Purdue University, West Lafayette, Indiana, USA.
- [3] Heacox, L. 2012. State of precision ag: data masters. *CropLife*, June 1, 2012.
- [4] ISO. 2007. Tractors and machinery for agriculture and forestry -- Serial control and communications data network -- Part 1: General standard for mobile data communication. ISO 11783-1:2007. Geneva, Switzerland: International Organization for Standardization (ISO).
- [5] Stone, M.L., K.D. McKee, C.W. Formwalt, and R.K. Benneweis. 1999. ISO 11783: an electronic communications protocol for agricultural equipment. ASAE Publication Number 913C1798. St. Joseph, Michigan, USA: ASABE.



## RESEARCH ARTICLE

# Effective sensor deployment based on field information coverage in precision agriculture

Wei An<sup>1</sup>, Song Ci<sup>1,2\*</sup>, Haiyan Luo<sup>3</sup>, Dalei Wu<sup>4</sup>, Viacheslav Adamchuk<sup>2</sup>, Hamid Sharif<sup>2</sup>, Xueyi Wang<sup>2</sup> and Hui Tang<sup>1</sup>

<sup>1</sup> High Performance Network Lab, IOA, Chinese Academy of Sciences, Beijing 100190, China

<sup>2</sup> Department of CEEN, University of Nebraska-Lincoln, Omaha, NE 68106, U.S.A.

<sup>3</sup> Cisco Systems, Inc, Austin, TX 78731, U.S.A.

<sup>4</sup> Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

## ABSTRACT

Coverage is an importance issue in wireless sensor networks. In this work, we first propose a novel notion of information coverage, which refers to the coverage efficiency of field information covered by deployed sensor nodes. On the basis of information coverage, we consider an optimization problem of how to partition the given field into multiple parcels and to deploy sensor nodes in some selected parcels such that the field information covered by the deployed sensor nodes meets the requirement. First, we develop two effective polynomial-time algorithms to determine the deployed locations of source nodes for information 1-coverage and  $q$ -coverage of the field, respectively, without consideration of communication, where information  $q$ -coverage implies that the field information in terms of information point is covered by at least  $q$  source nodes. Also, we prove the upper bound in the theoretical for the approximate solution derived by our proposed method. Second, another polynomial-time algorithm is presented for deriving the deployed locations of relay nodes. In the theoretical, this proposed algorithm can achieve the minimized number of relay nodes. Further, the related information 1-coverage algorithms are applied in our wireless sensor network-based automatic irrigation project in precision agriculture. Experimental results show the major trade-offs of impact factors in sensor deployment and significant performance improvements achieved by our proposed method. Copyright © 2013 John Wiley & Sons, Ltd.

## KEYWORDS

sensor deployment; information coverage; precision agriculture; communications; linear programming

## \*Correspondence

Song Ci, Department of CEEN, University of Nebraska-Lincoln, Omaha, NE 68106, U.S.A.; High Performance Network Lab, IOA, Chinese Academy of Sciences, Beijing 100190, P.R. China.

E-mail: sci@engr.unl.edu

## 1. INTRODUCTION

Wireless sensor networks (WSNs) consisting of a number of sensor nodes each capable of sensing, processing, and transmitting environmental information have attracted a lot of research attentions [1,2]. Because of the wide-range applications of WSNs, numerous research works have been carried out on sensor deployment with different objectives, such as extending the overall network lifetime by increasing energy efficiency [3–5] and improving field physical coverage [5–13]. However, little work that has been conducted on physical coverage addresses the related issues of the coverage of field information, that is, the apparent soil electrical conductivity [14], which usually describes the significant characteristics residing in the field in precision

agriculture. Therefore, how to deploy sensor nodes for providing the required physical coverage of field information is the main issue that will be addressed in this paper.

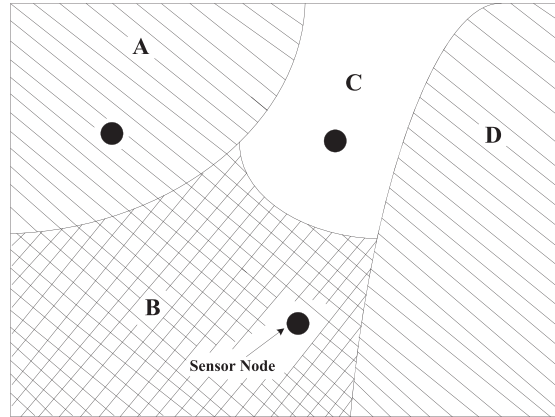
In this paper, we study a WSN-based automatic irrigation system in precision agriculture. In contrast to conventional agriculture that is practiced for the uniform application of fertilizer, herbicide, insecticides, fungicides, and irrigation, precision agriculture is a farming management concept based on observing and responding to intra-field variations [15–17]. In a piece of given agricultural land, the efficiency of a center-pivot irrigation system depends on its ability to meet the water demands of the crops that are growing [18]. Although limited water supplies can reduce crop yield, excessive irrigation leads to resource waste and even decrease of agricultural output.

WSN provides a useful tool for monitoring the soil spatial variability in order to assist decision making for field irrigation.

Effective sensor deployment, which affects not only the overall network performance but also the cost of the whole network, has gained increasing research interests in many aspects, such as improving field area coverage [6–8,19], extending the overall network lifetime [3–5]. In the literature, sensor deployment strategy is usually based on the concept of *physical coverage* [1], where the sensing range of a sensor is assumed to be fixed and the coverage range of the sensor is the area centered at itself with a radius equal to the sensing range. Any location is said to be covered if and only if it is within the sensing range of a sensor. However, different physical locations within a small piece of farming land may possess almost the same *field information*, which can be utilized to determine the occurrence frequencies of events, for example, water irrigation. Here, field information can be diverse, including the apparent soil electrical conductivity, elevation, soil profile, soil salinity and sodicity, and water storage capacity [14,20–22]. Consequently, to monitor only a portion of locations that correspond to the full range of field information is sufficient to monitor the whole field. Therefore, monitoring the events associated with the field information provides a new perspective on sensor deployment.

In our research project in precision agriculture, field information is the water storage capacity of the locations, which is closely associated with the apparent soil electrical conductivity (ECa). Because the field information changes spatially, it is necessary to deploy some sensor nodes such that the water content in the soil can be monitored and the optimum quantity of irrigation water can be derived. The soil temperature and moisture sensor nodes are deployed in the selected locations to collect the field information. This is *information coverage*, which will be discussed in detail in Section 3. Because multiple locations share the same value of water storage capacity, sensor deployment based on information coverage can remarkably decrease the number of required sensor nodes. Consider a simple example shown in Figure 1, areas *A* and *D* share the same field information. This implies that they consume the water at the same speed. Let the initial water content of two areas be identical. Then, one single sensor deployed in *A* or *D* is sufficient to collect the water consumption information for both areas. Similarly, deploying one sensor in each area of *B* and *C* can also sufficiently monitor the water scarcity of these two areas.

In this paper, on the basis of our research project on the WSN-based irrigation system in precision agriculture, we investigate the optimization of information coverage for sensor deployment. The main contributions of this paper are as follows: (i) in contrast to conventional physical coverage, we propose a novel notion of *information coverage*, which can remarkably decrease the number of deployed sensor nodes for monitoring temporal field information; (ii) the major factors influencing sensor deployment based on information coverage are theoretically



**Figure 1.** Illustration of information coverage, where different physical locations can share the same field information.

analyzed; (iii) three polynomial-time algorithms are presented to achieve the approximate sensor deployment scheme, and the performance of these algorithms are also studied in the theoretical; and (iv) we apply the proposed information

1-coverage algorithms in our WSN-based automatic irrigation system of precision agriculture and achieve significant performance improvement.

The rest of the paper is organized as follows. Section 2 reviews the related work. Section 3 discusses information coverage and the major factors that influence sensor deployment, and then formulate the sensor deployment problem on the basis of information coverage. Section 4 develops three algorithms. Further, in Section 5, we discuss how to use the proposed method in our WSN-based soil monitoring project in precision agriculture. Finally, we conclude the paper in Section 6.

## 2. RELATED WORK

Various effective approaches have been adopted to maximize the physical coverage of deployed sensor nodes. By representing the sensing field as a grid of points, Charkrabarty *et al.* [9] exploited a grid coverage strategy for surveillance with distributed sensor networks and utilized an integer linear programming to decrease the cost, expended on sensor nodes, for the complete coverage of a given area. On the basis of the principle of moving sensor nodes from densely deployed areas to sparsely deployed areas, Wang *et al.* [10] designed and evaluated three distributed self-deployment protocols for mobile sensor nodes. After discovering a coverage hole, the proposed protocols can dynamically calculate the target positions of the sensor nodes where they should move to. To provide the desired coverage of a target point set, Xu *et al.* [11] developed an integer linear programming formulation to achieve the minimized deployment cost and proposed a greedy heuristic algorithm to solve the problem. Further, to reduce

the number of sensor nodes needed to provide guaranteed barrier coverage, Yang *et al.* [5] proposed multi-round sensor deployment that splits sensor deployment into multiple rounds and solves placement errors that often accompany sensor deployment. Considering delivery failures occur constantly in ZigBee wireless applications because of node movements and network topology changes, Shih *et al.* [23] proposed a scheme that exploits the regularity to improve the data delivery ratio in ZigBee WSNs, which deploys the network nodes and constructs the tree topology by using the mobility regularity imposed by the physical environment. Wang *et al.* [24] studied coverage in mobile and heterogeneous WSNs; specifically, they investigated asymptotic coverage under uniform deployment model with independent and identically distributed and 1D random walk mobility model, respectively. Mahboubi *et al.* [25] presented efficient sensor deployment algorithms to increase coverage in a mobile sensor network with a prescribed priority assignment for different points in the sensing field. Besides those works on 2D ideal plane coverage, Kong *et al.* [26] expanded their research vision from 2D ideal plane coverage and 3D full space coverage to surface coverage.

Aside from the aforementioned works, WSNs have been increasingly used in precision agriculture. In order to optimize available resources and to increase the yield, Martinelli *et al.* [27] investigated the deployment of sensor nodes for collecting the data of apple trees in a real-time manner by using spatially distributed sensor nodes. Also, to acquire time-variant environmental information in a real-time fashion, Xiao *et al.* [15] discussed hardware and software design, network topology, and network communication protocol. Further, our previous work [14] provided a simple analytical methodology for assessing the monitoring quality of a set of strategic deployed locations of sensor nodes based on the information of apparent soil electrical conductivity and field elevation. However, in that work, we divided the field into nine big blocks and selected the optimal deployed locations in each block. Considering the complexity of sensor deployment that method is preliminary and heuristic, which lacks the effective selection of strategic deployed locations with comprehensive consideration of the overall field information, the theoretical analysis of the performance of the proposed method and the major trade-offs of impact factors such as communication radius under the blockage from the land surface crop density and the parcel size of the monitoring area for each sensor for the sake of the irrigation management in precision agriculture.

### 3. PROBLEM STATEMENTS

In a given land, the field information value  $\tilde{y}$  can be generally represented as

$$\tilde{y} = f(x_1, x_2, x_3) \quad (1)$$

where  $x_1$  and  $x_2$  are the horizontal and vertical coordinates, respectively, and  $x_3$  is the elevation of the location. For instance, as discussed in our previous work [14], the field information (i.e., water storage capacity) is closely associated with the ECa of the locations in the field. Thus,  $\tilde{y}$  may represent the ECa value of location  $(x_1, x_2)$ . Given a crop field, the field usually contains considerable different information values. For these information values, we give the definition as follows.

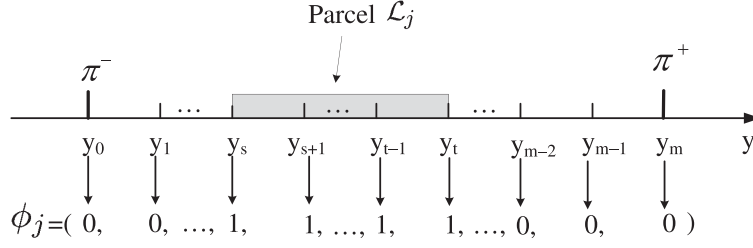
**Definition 1** (Information Coverage). *A field information value is said to be covered if at least one location associated with this value is monitored by one sensor node. The field information is completely covered, namely complete information coverage, if all the information values are monitored by the deployed sensor nodes. The field is called information  $q$ -coverage, if each information value is monitored by at least  $q$  sensor nodes at a certain location in the field.*

*It is worth noting that the proposed notion of information coverage is significantly different from conventional physical coverage because of the following reasons. First, complete field information coverage usually needs far fewer sensor nodes than complete physical coverage; for example, areas  $A$  and  $D$  can be completely covered by one sensor node as illustrated in Figure 1, because multiple physical locations usually share almost the same field information values. Thus, information coverage is preferred from the cost-effective perspective. Second, for information  $q$ -coverage, an information value is covered by  $q$  sensor nodes each of which may be deployed in any location of the field, whereas  $q$ -physical coverage requires all  $q$  sensor nodes deployed in the vicinity of the covered location.*

#### 3.1. Field information discretization

Before going further, we first adopt a discretization strategy for tackling numerous field information values. Without losing generality, assume  $[\pi^-, \pi^+]$  is the range of values for field information  $y$ . Because the information values are usually very complex and uneven, it is not realistic to use them directly. Thus, we process them by discretizing the range  $[\pi^-, \pi^+]$  at step size  $\tau$ ; that is, as shown in Figure 2,  $[\pi^-, \pi^+]$  can be uniformly discretized as  $y_0 = \pi^- < y_1 < y_2 < \dots < y_m = \pi^+$ , where  $y_i = y_0 + i \cdot \tau$ ,  $i = 1, 2, \dots, m$ . For the sake of description,  $y_i$  is called *information point*. As for field information, the given field can be uniformly sampled to achieve a great number of locations with the related data, represented by  $\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_r$ , where  $r$  is the number of information points appears in the field. For each information point  $y_i$ , it namely appears in the field if and only if there exists one information value  $\tilde{y}_{i_0}$  satisfying  $|y_i - \tilde{y}_{i_0}| < \frac{1}{2}\tau$ .

On the basis of the discretization, a column vector  $\phi_j$  for parcel  $\mathcal{L}_j$  can be formed as  $\phi_j^T = (\phi_{j1}, \phi_{j2}, \dots, \phi_{jm})$ , where  $\phi_j^T$  denotes the transpose of  $\phi_j$  and  $\phi_{ji} = 1$  if



**Figure 2.** Discretizing the range  $[\pi^-, \pi^+]$  of field information values to derive the corresponding vector  $\phi_j$  for parcel  $\mathcal{L}_j$ .

information point  $y_i$  is in  $\mathcal{L}_j$  and otherwise  $\phi_{ji} = 0$ . Then, the vectors  $\phi_1, \phi_2, \dots, \phi_n$  of all the parcels form a matrix  $\Phi = (\phi_1, \phi_2, \dots, \phi_n)$ .

For step size  $\tau$ , if it is big enough, many vectors derived from the parcels will be the same. This seriously affects the accuracy of sensor deployment because the potential differences between any two parcels cannot be reflected from the derived vectors. Here, we discuss the relations among the parcels as follows. As for two parcels  $\mathcal{L}_j$  and  $\mathcal{L}_k$ , we have the following three relations:

- Relation i:  $[\pi_j^-, \pi_j^+] \cap [\pi_k^-, \pi_k^+] = \emptyset$ ;
- Relation ii:  $[\pi_j^-, \pi_j^+] \subseteq [\pi_k^-, \pi_k^+]$  or  $[\pi_j^-, \pi_j^+] \supseteq [\pi_k^-, \pi_k^+]$ ;
- Relation iii:  $[\pi_j^-, \pi_j^+] \cap [\pi_k^-, \pi_k^+] \neq \emptyset$  and  $[\pi_j^-, \pi_j^+] \not\subseteq [\pi_k^-, \pi_k^+]$  and  $[\pi_k^-, \pi_k^+] \not\subseteq [\pi_j^-, \pi_j^+]$ ;

where  $[\pi_j^-, \pi_j^+]$  and  $[\pi_k^-, \pi_k^+]$  are the ranges of field information points in parcels  $\mathcal{L}_j$  and  $\mathcal{L}_k$ , respectively. If the field information points of parcels  $\mathcal{L}_j$  and  $\mathcal{L}_k$  meet Relation i, their characteristic vectors  $\phi_j$  and  $\phi_k$  can be easily distinguished because they have no same information points. Further, if the information points of  $\mathcal{L}_j$  and  $\mathcal{L}_k$  meet Relation ii, we do not need to distinguish the two sets of information points because the bigger one is always covered by the smaller one. Therefore, we only need to consider Relation iii; that is, two sets of information points intersect but do not belong to each other.

**Proposition 1.** Suppose in the field, there are totally  $n$  parcels, denoted as  $\{\mathcal{L}_j\}_{j=1}^n$ , any pair of which meets Relation iii. Then the range of the information points needs to be divided into at least  $\log_2 n - 1$  intervals such that any two vectors  $\phi_j$  and  $\phi_k$  derived by parcels  $\mathcal{L}_j$  and  $\mathcal{L}_k$  are different.

**Proposition 2.** Suppose  $\phi_j$  and  $\phi_k$  are two vectors derived from parcels  $\mathcal{L}_j$  and  $\mathcal{L}_k$ , respectively, which satisfy Relation iii, then  $\phi_j$  and  $\phi_k$  are different if  $\tau \leq \max\{|\pi_j^- - \pi_k^-|, |\pi_j^+ - \pi_k^+|\}$ , where  $[\pi_j^-, \pi_j^+]$  and  $[\pi_k^-, \pi_k^+]$  are the ranges of information points for  $\mathcal{L}_j$  and  $\mathcal{L}_k$ , respectively.

Further, for any two parcels  $\mathcal{L}_j$  and  $\mathcal{L}_k$  ( $1 \leq j < k \leq n$ ) in  $\{\mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_n\}$  that satisfy Relation iii, we define

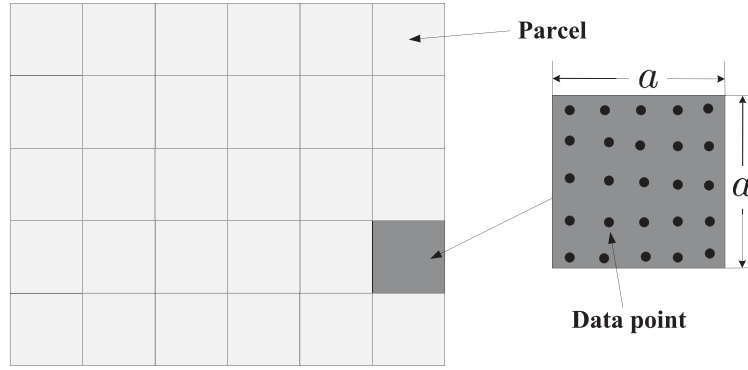
$$\tau_{jk} = \max\{|\pi_j^+ - \pi_k^+|, |\pi_j^- - \pi_k^-|\} \quad (2)$$

Then, to guarantee that any pair of vectors in  $\{\phi_1, \phi_2, \dots, \phi_n\}$  are different, the step size  $\tau$  should satisfy

$$\begin{aligned} \tau &\leq \min_{1 \leq j, k \leq n} \{\tau_{jk}\} \\ &= \min_{1 \leq j, k \leq n} \left\{ \max\{|\pi_j^+ - \pi_k^+|, |\pi_j^- - \pi_k^-|\} \right\} \quad (3) \end{aligned}$$

### 3.2. Field partition

For deploying sensor nodes, in this paper, we introduce a strategy to partition the given field into a number of same-size square parcels with side length  $a$ , as shown in Figure 3, because the same-size parcels are convenient for irrigation management in precision agriculture. Within each parcel, we deploy an array of soil temperature and moisture sensor nodes for collecting the water storage information and one relay node for relaying the collected data to the sink node. For the sake of description, we call all the nodes deployed in one parcel as a whole *parcel node*. It is worth noting that the number of sampled field information data (ECa) within the square parcels increases with the increase of  $a$ , which results in the decrease of the number of parcel nodes required for field information monitoring. However, this also leads to the decrease of the sensing accuracy and degrades the overall performance of the field monitoring because each parcel node covers too much information points in each parcel, which are usually fused as one proper information value for representing the field water storage information and sending to the sink node. On the other hand, when  $a$  gets small, the collected data will be more accurate, but the number of parcel nodes deployed for field monitoring will increase. This unavoidably results in the increase of the deployment cost. Therefore, how to determine the proper size of the parcels is a critical issue to achieve a good trade-off between monitoring quality and deployment cost.



**Figure 3.** Illustration of the field partition, where the given area is divided into many square parcels of the same size.

Let  $\lambda_j$  be the standard deviation of all the field information points  $\{y_i\}_{i=1}^p$  in parcel  $\mathcal{L}_j$ , that is,

$$\lambda_j = \sqrt{\sum_{y_i \in \mathcal{L}_j} [y_i - \bar{y}]^2} \quad (4)$$

where  $\bar{y} = \frac{1}{p} \sum_{y_i \in \mathcal{L}_j} y_i$ , because, as discussed earlier, the  $\lambda_j$  is significantly affected by  $a$ , and  $\lambda_j$  can be controlled by adjusting  $a$ . Now, to guarantee the monitoring accuracy of the field, a small real number  $\delta > 0$  is given as the threshold of the  $\lambda_j$ . Then, the proper value  $a_j^*$  can be derived by

$$a_j^* = \arg \max \{a_j : a_j \text{ satisfies } \lambda_j < \delta\} \quad (5)$$

For all the parcels and given  $\delta$ , the proper  $a^*$  can be calculated as

$$a^* = \min_{1 \leq j \leq n} \{a_j^*\} \quad (6)$$

where  $n$  is the total number of parcels.

### 3.3. Communication considerations

In general, the communication prerequisite for any two nodes is that the Euclid distance between them is smaller than communication radius  $R$ . Suppose  $A$  and  $B$  are two nodes with coordinates  $(x_1^A, x_2^A, x_3^A)$  and  $(x_1^B, x_2^B, x_3^B)$ , respectively, the distance between them can be calculated as

$$d(A, B) = \sqrt{\sum_{i=1}^3 (x_i^A - x_i^B)^2} \leq R \quad (7)$$

Aside from distance, however, the communication between nodes  $A$  and  $B$  is also significantly affected by the blockage from the field surface and high density crops. For the blockage from the field surface, we first give a reasonable

assumption that any two nodes can communicate with each other only when their line-of-sight does not intersect with the field surface. Then, the blockage from the field surface can be known by judging whether the segment between two relay nodes intersects with the field surface. As for the blockage from crops, extensive on-site tests are conducted in the field to derive the real communication radius of the relay nodes, which is usually much smaller than advertised. The detail of the on-site communication radius tests can be found in Section 5.

In this paper, we do not consider the interference among nodes because (i) the number of deployed nodes is usually far less than the number of parcels; (ii) the event of data collection does not happen frequently; and (iii) the traffic-adaptive medium access protocol reduces the interference among nodes by allowing nodes to switch to a low-power, idle state whenever they are not transmitting or receiving [28].

### 3.4. Problem formulation

For a given field, on the basis of the aforementioned discretization, the appearance times of information point  $y_1, y_2, \dots, y_m$  in the whole field can be calculated as

$$\begin{aligned} \psi_0 &= \sum_{j=1}^n \phi_j \\ &= \left( \sum_{j=1}^n \phi_{j1}, \sum_{j=1}^n \phi_{j2}, \dots, \sum_{j=1}^n \phi_{ji}, \dots, \sum_{j=1}^n \phi_{jm} \right)^T \end{aligned} \quad (8)$$

where the appearance times of  $y_i$  in the field is represented by the  $i$ th component  $\psi_0^{(i)}$  of  $\psi_0$  with

$$\psi_0^{(i)} = \sum_{j=1}^n \phi_{ji} \quad (9)$$

For simplicity, in this work, we assume the field is divided into  $n$  same-size parcels. A set of 2-tuple  $\{(\mathcal{L}_j, \mathcal{N}_j)\}_{j=1}^n$



is adopted to represent the sensor deployment scheme  $\mathcal{S}$ , where 2-tuple  $(\mathcal{L}_j, \mathcal{N}_j)$  implies parcel node and  $\mathcal{N}_j$  is deployed in parcel  $\mathcal{L}_j$ . We define vector  $\mathbf{c}_S$  for the scheme  $\mathcal{S}$  such that  $c_j = 1$  if  $\mathcal{N}_j$  is deployed in parcel  $\mathcal{L}_j$ ; otherwise  $c_j = 0$ . Further, the appearance times of  $y_1, y_2, \dots, y_m$  in the scheme  $\mathcal{S}$  can be calculated as

$$\psi_S = \Phi \cdot \mathbf{c}_S = \left( \sum_{j=1}^{\tau} \phi_{j1}, \sum_{j=1}^{\tau} \phi_{j2}, \dots, \sum_{j=1}^{\tau} \phi_{jm} \right)^T \quad (10)$$

On the basis of Equations (8) and (10), we define the monitoring efficiency as follows.

**Definition 3 (Monitoring Efficiency).** The monitoring efficiency  $\eta$  for information  $q$ -coverage is defined as

$$\eta(\mathbf{c}_S, q) = \frac{\sum_{i=1}^m h_q(\psi_S^{(i)})}{\sum_{i=1}^m h_q(\psi_0^{(i)})} \quad (11)$$

where  $\psi^{(i)}$  denotes the  $i$ th component of  $\psi$ .  $h_q(\psi^{(i)})$  is a function with  $h_q(\psi^{(i)}) = 1$  if  $\psi^{(i)} \geq q$ , otherwise  $h_q(\psi^{(i)}) = 0$ .

With this definition, monitoring efficiency  $\eta$  is the ratio between the number of information points covered  $q$  times by  $\mathcal{S}$  and the total number of information points in the field with  $0 \leq \eta(\mathbf{c}_S, q) \leq 1$ . As for connectivity, it is required to guarantee that any parcel node  $\mathcal{N}_j$  ( $1 \leq j \leq |\mathcal{S}|$ ) in scheme  $\mathcal{S}$  can transmit the collected data to sink node  $\mathcal{N}_{SN}$ ; that is, any node is connected to the sink node, denoted by  $\mathcal{N}_j \leftrightarrow \mathcal{N}_{SN}$ .

Let  $\epsilon > 0$  be a small real number. Our optimization problem is to derive the sensor deployment scheme  $\mathcal{S}$  with the minimum number of parcels such that it reaches the given monitoring efficiency and maintains the connectivity between parcel nodes and sink node  $\mathcal{N}_{SN}$ , that is,

$$\min_{\mathbf{c}_S \in \{0,1\}^n} \|\mathbf{c}_S\|_1 \quad (12)$$

$$\text{Subject to: } \begin{cases} \eta(\mathbf{c}_S, q) \geq 1 - \epsilon \\ \mathcal{N}_j \leftrightarrow \mathcal{N}_{SN} \text{ for } \forall c_j \neq 0 \end{cases}$$

where  $\|\mathbf{c}_S\|_1 = \sum_{i=1}^n |c_i|$ . This problem is an integer programming problem, which is classified as non-deterministic polynomial-time (NP)-hard and also one of Karp's 21 NP-complete problems [29].

## 4. INFORMATION-COVERAGE-BASED SENSOR DEPLOYMENT

In our research project, sensor nodes include two types of nodes: relay nodes and parcel nodes, which are much more expensive than relay nodes because it includes an array of temperature and moisture sensor nodes. Jointly considering

cost and complexity, we first achieve the approximate sensor deployment scheme without consideration of communication on the basis of the *Set Covering Theory* [29,30], and we next develop an optimal algorithm for deploying additional relay nodes for the connectivity of the network.

### 4.1. Finding approximate sensor deployment scheme

#### 4.1.1. Information 1-coverage.

To derive the information 1-coverage sensor deployment scheme of the formulated problem (Equation (12)), we give the following definition.

**Definition 4 (Minus Operation).** The minus operation is defined for vectors  $\phi_j$  and  $\phi_k$  as

$$\phi_j \ominus \phi_k = \begin{pmatrix} \phi_{j1} \ominus \phi_{k1} \\ \phi_{j2} \ominus \phi_{k2} \\ \vdots \\ \phi_{jm} \ominus \phi_{km} \end{pmatrix} \quad (13)$$

where, for  $1 \leq l \leq m$ ,

$$\phi_{jl} \ominus \phi_{kl} = \begin{cases} \phi_{jl} - \phi_{kl} & \text{if } \phi_{jl} > \phi_{kl} \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

This minus operation aims to eliminate the information points included in parcel  $\mathcal{L}_k$  from parcel  $\mathcal{L}_j$ . Let  $\mathbf{b}$  be an  $m$ -dimensional vector such that  $b_i = 1$  if  $y_i$  is included in the given field, otherwise  $b_i = 0$ . Considering the NP-hardness of problem Equation(12), a greedy strategy is exploited to select the parcel containing the largest number of uncovered information points of  $\{y_i\}_{i=0}^m$  at each step. Then, we develop Algorithm 1 to derive the information 1-coverage sensor deployment scheme.

---

**Algorithm 1** Deriving the information 1-coverage scheme

---

- 1:  $\epsilon > 0$  and  $\mathcal{S} = \emptyset$
  - 2:  $I = \{j\}_{j=1}^n$  and  $\Lambda = \emptyset$  /\* index set\*/
  - 3:  $\mathbf{b}^T = (1, 1, \dots, 1)$ ;
  - 4: **while**  $\eta(\mathbf{c}_S, 1) < 1 - \epsilon$  **do**
  - 5: (a) Select parcel  $\mathcal{L}_j$  from  $\{\mathcal{L}_k : k \in I \setminus \Lambda\}$ , whose vector  $\phi_i$  minimizes  $\{\|\mathbf{b} \ominus \phi_j\|_1 : j \in I \setminus \Lambda\}$ .
  - (b) Let  $\mathcal{S} = \mathcal{S} \cup \{(\mathcal{L}_j, \mathcal{N}_j)\}$  and  $\Lambda = \Lambda \cup \{j\}$ ;
  - (c)  $\mathbf{b} = \mathbf{b} \ominus \phi_j$ ;
  - 6: **end while**
  - 7: **return**  $\mathcal{S}$
- 

In Algorithm 1, selecting parcel  $\mathcal{L}_j$  from  $\{\mathcal{L}_k : k \in I \setminus \Lambda\}$  takes  $O(n)$  time, where  $n$  is the number of the parcels and where  $\Lambda$  is the set of indices of the selected parcels. We select the parcel  $\mathcal{L}_j$  at most  $n$  times. Therefore, the computational complexity of Algorithm 1 is  $O(n^2)$ .

As for the theoretical performance, motivated by [31], the following theorem is given.

**Proposition 5.** *The sensor deployment scheme found by Algorithm 1 is a  $\ln \frac{m}{|OPT|}$ -approximation to OPT, where OPT and |OPT| are the optimal sensor deployment scheme and the number of nodes in OPT.*

#### 4.1.2. Information $q$ -coverage.

For efficient field monitoring, a strategy for information  $q$ -coverage is adopted to cover each information point with at least  $q$  parcel nodes. Information  $q$ -coverage can provide fault tolerance in that it is able to monitor all the field information as long as no more than  $q - 1$  parcel nodes fail. To implement information  $q$ -coverage, one simple way is to deploy  $q$  parcel nodes in each selected parcel derived by Algorithm 1. However, deploying  $q$  parcel nodes in one parcel usually may not effectively improve the fault tolerance of the field information coverage. Therefore, for effective information  $q$ -coverage, one best way is to deploy  $q$  parcel nodes in different parcels. With this aim, we achieve the information  $q$ -coverage with Algorithm 2 adapted from Algorithm 1.

---

**Algorithm 2** Deriving the information  $q$ -coverage scheme

---

- 1:  $\epsilon > 0$  and  $\mathcal{S} = \emptyset$
  - 2:  $I = \{j\}_{j=1}^n$  and  $\Lambda = \emptyset$
  - 3:  $\mathbf{b}^T = (q, q, \dots, q)$ ;
  - 4: **while**  $\eta(\mathbf{c}_S, q) < 1 - \epsilon$  **do**
  - 5: (a) Select parcel  $\mathcal{L}_j$  from  $\{\mathcal{L}_j : j \in I \setminus \Lambda\}$ , whose vector  $\phi_j$  minimizes  $\{\|\mathbf{b} \ominus \phi_j\|_1 : j \in I \setminus \Lambda\}$ .
  - (b) Let  $\mathcal{S} = \mathcal{S} \cup \{\mathcal{L}_j, \mathcal{N}_j\}$  and  $\Lambda = \Lambda \cup \{j\}$ .
  - (c)  $\mathbf{b} = \mathbf{b} \ominus \phi_j$
  - 6: **end while**
  - 7: **return**  $\mathcal{S}$
- 

As for computational complexity of Algorithm 2, because the computational complexity of Algorithm 1 is  $O(n^2)$ , the computational complexity of Algorithm 2 is  $O(n^2)$ , where  $n$  is the number of parcels.

Because Algorithm 1 is repeated  $q$  times, each information point is covered by at least  $q$  parcels. Thus, one necessary condition for the existence of sensor deployment scheme  $\mathcal{S}$  is that for any information point, we can find  $q$  parcels, which cover the information point. Because of the limited number of parcels, the number  $q$  must have an upper bound. The literature [32,33] investigated the set  $q$ -cover for the physical coverage in the field monitoring. Provided that one parcel node is deployed in each parcel, the algorithms proposed in [32] can also be utilized to derive the upper bound of  $q$ .

## 4.2. Deploying relay nodes

As for the resulting scheme  $\mathcal{S}$  derived with Algorithms 1 and 2, if scheme  $\mathcal{S}$  is connected, that is, for any  $\mathcal{N}_j \in \mathcal{S}$ ,

we have  $\mathcal{N}_j \leftrightarrow \mathcal{N}_{SN}$ , then we do not need to deploy additional relay nodes for relaying the collected data to the sink node  $\mathcal{N}_{SN}$ . Usually, scheme  $\mathcal{S}$  is composed of  $\nu$  connected components, as  $\{\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_\nu\}$ . Here, the connected component is a subnetwork in which any two nodes are connected with each other and which is connected with the other nodes not belonging to it. Especially, a single parcel node is also viewed as a connected component. To guarantee the data transmission from any parcel node in  $\mathcal{S}$  to the sink node, it is required to make sure that  $\{\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_\nu\}$  and  $\mathcal{N}_{SN}$  are included in one connected component by deploying additional communication or relay nodes.

Before going further, the distance between two connected components  $\mathcal{C}_j$  and  $\mathcal{C}_k$  is defined as

$$d(\mathcal{C}_j, \mathcal{C}_k) = \min_{\forall \mathcal{N}_j \in \mathcal{C}_j, \mathcal{N}_k \in \mathcal{C}_k} \sqrt{\sum_{i=1}^3 (x_i^{\mathcal{N}_j} - x_i^{\mathcal{N}_k})^2} \quad (15)$$

where  $(x_1^{\mathcal{N}_j}, x_2^{\mathcal{N}_j}, x_3^{\mathcal{N}_j})$  and  $(x_1^{\mathcal{N}_k}, x_2^{\mathcal{N}_k}, x_3^{\mathcal{N}_k})$  are the deployed locations of parcel node  $\mathcal{N}_j$  and  $\mathcal{N}_k$ , respectively. In our project, the sink node  $\mathcal{N}_{SN}$  is assumed to be deployed at the specific place, denoted as connected component  $\mathcal{C}_0$ . Let  $\mathcal{S} = \{\mathcal{C}_j\}_{j=0}^\nu$ . Now Algorithm 3 is developed for deploying additional relay nodes such that each parcel node can transmit the collected data to the sink node  $\mathcal{N}_{SN}$ .

---

**Algorithm 3** Deploying additional relay nodes.

---

- 1: **while** There exists a component  $\widehat{\mathcal{C}}$  disconnected to  $\mathcal{N}_{SN}$  **do**
  - 2: Calculate the  $d$  among all the connected components in  $\mathcal{S}$  and select  $\mathcal{C}_j$  and  $\mathcal{C}_k$  with the smallest  $d$ , denoted as  $d_{\min}$ . Let  $d_{\min}$  is achieved at  $\mathcal{N}_{j0} \in \mathcal{C}_j$  and  $\mathcal{N}_{k0} \in \mathcal{C}_k$ .
  - 3: Let  $\mathcal{S} = \mathcal{S} \setminus \{\mathcal{C}_j, \mathcal{C}_k\}$ .
  - 4: Deploy  $\lceil d_{\min}/R \rceil$  relay nodes to connect  $\mathcal{N}_{j0}$  and  $\mathcal{N}_{k0}$ , where  $R$  is the communication radius. Then,  $\mathcal{C}_j, \mathcal{C}_k$  and the added  $\lceil d_{\min}/R \rceil$  relay nodes constitute a new connected component  $\mathcal{C}'_j$ .
  - 5: Let  $\mathcal{S} = \mathcal{S} \cup \{\mathcal{C}'_j\}$ .
  - 6: **end while**
  - 7: **return**  $\mathcal{S}$
- 

In Algorithm 3, calculating the  $d$  among all the connected components and selecting  $\mathcal{C}_j$  and  $\mathcal{C}_k$  with the smallest  $d_{\min}$  take the  $O(\nu^2)$  time, where  $\nu$  is the number of the connected components. We repeat the aforementioned procedure at most  $\nu - 1$  times. Therefore, the computational complexity of Algorithm 3 is  $O(\nu^3)$ .

Noting that in scheme  $\mathcal{S}$  derived from Algorithm 3, all sensor nodes in each connected component can deliver the collected data to the sink node  $\mathcal{N}_{SN}$ . Here, this algorithm is motivated by the idea of *Prim's minimum spanning*

*tree algorithm* in [34]. The theoretical performance of this algorithm is presented as follows:

**Proposition 6.** *Algorithm 3 achieves the least number of relay nodes deployed for connecting the connected components.*

**Remarks 1.** *Because of the complexity of the formulated problem, more efforts are required for improving the information  $1/q$ -coverage because of the limitation of the proposed algorithms.*

- (1) *As for field information discretization, too much field information points pose a great challenge for finding the  $1$ -coverage or  $q$ -coverage scheme because it requires the huge capability of storing space and computation for the huge space of field information points.*
- (2) *Because the formulated problem is NP-hard, the proposed algorithms for information  $1/q$ -coverage of the field are approximated ones. More efforts are needed to develop more efficient algorithms for set covering problem and information coverage problem.*

## 5. SENSOR DEPLOYMENT FOR AUTOMATIC IRRIGATION SYSTEM BASED ON INFORMATION COVERAGE

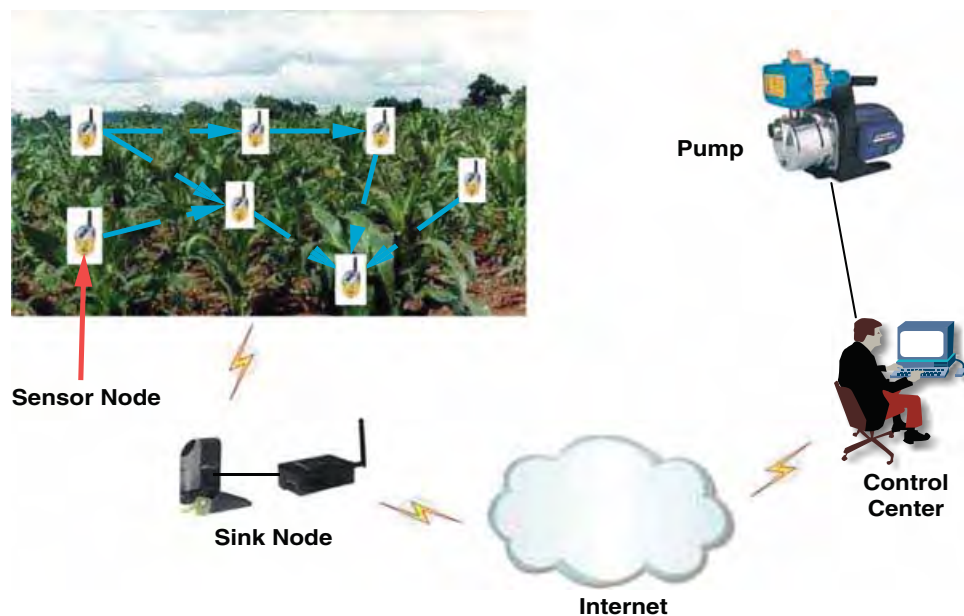
In this section, we utilize our proposed method in the WSN-based automatic irrigation system in precision agriculture and evaluate the achieved significant performance.

### 5.1. Field data collecting

Figure 4 is the overall architecture of the WSN-based irrigation system used in our research project, where sensor nodes, including the source and relay nodes, constitute the WSN in a self-organizing fashion and send the collected data to the sink node. The sink node aggregates the collected data and further relays them to the users via the Internet. Then the remote control center processes the collected data and makes decisions on whether or not to send a command to the farming field for starting the irrigation system and what the quantity of water is.

To obtain high resolution maps of ECa, a Veris® 3150 unit (Mobile Sensor Platform, Veris Technologies, Inc., Salina, Kansas) equipped with an RTK-level AgGPS® 442 GNSS receiver (Trimble Navigation Limited, Sunnyvale, California, USA) is used to map a 37-ha field located at the University of Nebraska-Lincoln Agricultural Research and Development Centre near Mead, Nebraska, USA [14]. The shallow measurements of ECa, that is, 0–30 cm, are exploited in the research project. Both ECa and elevation data are collected with 1-Hz mapping frequency while moving at approximately 1.5 m/s travel speed with a 13.7-m swath width, which resulted in about 30 000 data points in the given land.

As shown in Figure 5, the eKo eN 2100 sensor nodes are used as the relay nodes to relay the data over the WSN, and the eKo eS 1101 watermark sensor nodes are used to monitor soil moisture and temperature [35]. Considering the interference and the blockage from field surface and crops, we test the real communication radii of eN 2100 sensor nodes in the field. Because the sprinkler system for field irrigation is 1.5 feet above the canopy, as



**Figure 4.** The overall architecture of the wireless sensor network-based automatic irrigation system used in precision agriculture.

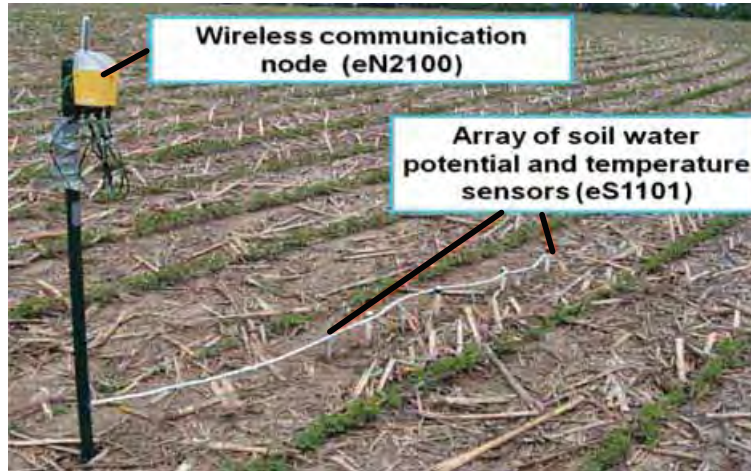


Figure 5. The communication node (i.e., relay node) and the array of soil water potential and temperature sensor nodes.

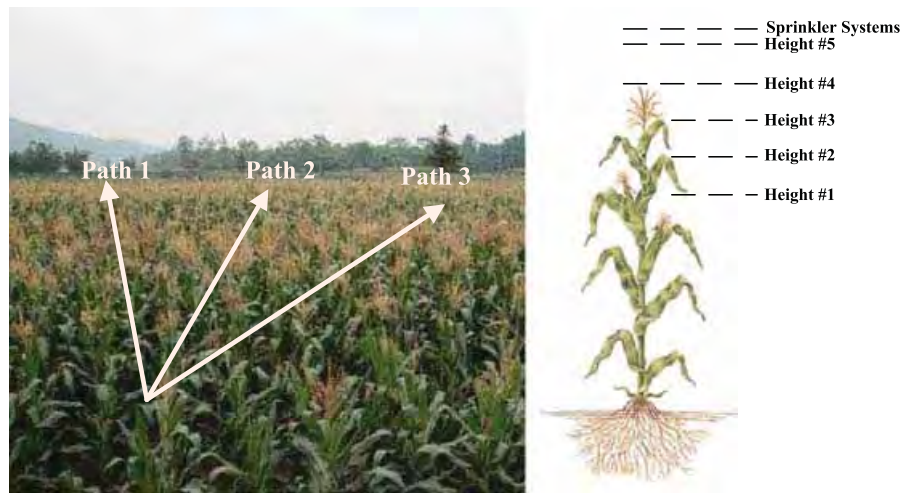


Figure 6. Three paths and five heights for testing the wireless communication radii in the field.

shown in Figure 6, we select five relative heights compared with the canopy:  $-3$ ,  $-2$ ,  $-1$ ,  $0$ , and  $1$  feet, where “ $-$ ” denotes that the height is lower than the canopy. Also, as to the influence from the density of crops, we select three paths to test the communication distance among relay nodes. Table I shows the results of the tested communication radii.

Table I. The communication distances are tested in the corn field, where the unit of communication distance is meter.

Height #	1	2	3	4	5
Path 1	0–47.55	0–85.95	0–91.44	0–182.88	0–273.10
Path 2	0–27.43	0–58.52	0–128.93	0–187.45	0–283.46
Path 3	0–26.52	0–77.72	0–99.67	0–111.56	0–365.76

## 5.2. The major parameters influencing sensor deployment

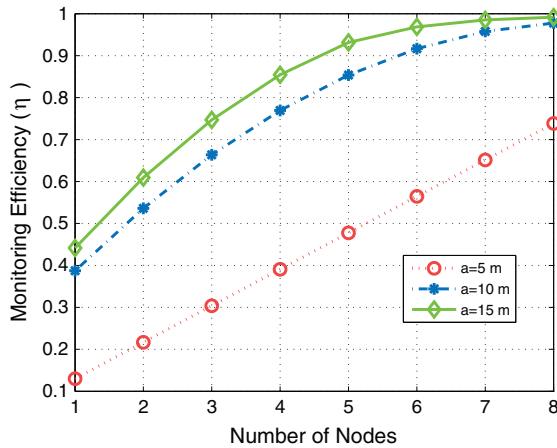
### 5.2.1. Impact of parcel size.

To understand the influence of parcel size on field monitoring, three different lengths,  $a = 5, 10$ , and  $15$  m, are selected to show the relations between monitoring efficiency ( $\eta$ ) and the number of deployed parcel nodes.

In Figure 7, it is observed that the sensor deployment scheme achieves better monitoring efficiency when the

field is partitioned with larger parcel size. This is reasonable because the parcel of large size contains more information points than that of small size. However, because the entire parcel is usually viewed as a unit for the sake of irrigation management, the large parcel size will degrade the control accuracy because of the low accuracy of the collected data. Therefore, the proper parcel size is necessary to be derived for accurate irrigation management of the field.

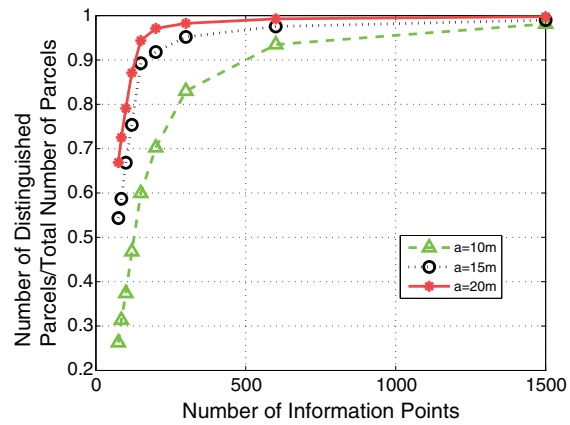




**Figure 7.** Comparison of monitoring efficiency versus the number of sensor nodes under different parcel sizes.

**5.2.2. Impact of discretization.**

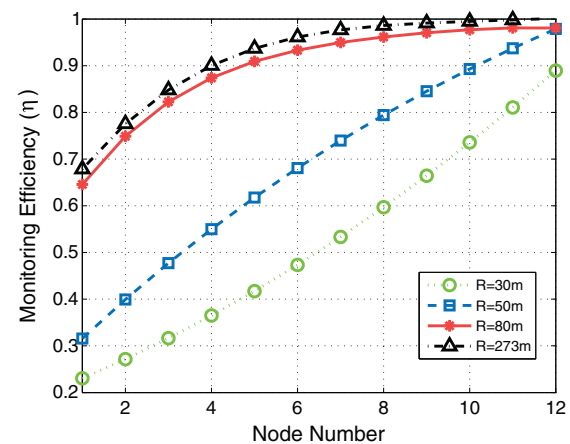
To show the influence of discretization, we divide the range [0.23, 14.56] of the storage capacity of the field into 75, 85, 100, 120, 150, 200, 300, 600, and 1500 information points with step sizes of 0.20, 0.175, 0.15, 0.125, 0.10, 0.075, 0.05, 0.025, and 0.01, respectively. In Figure 8, it is observed that the long step is required for the big parcel to distinguish different parcel vectors, whereas the short step is appropriate. This is reasonable because the difference between two big parcels is usually larger than that between two small parcels. Because the parcel size is determined by the monitoring accuracy, an analyzing method is required to derive the proper step size with consideration of all the parcels as discussed in Section 3.1.



**Figure 8.** Comparison of the ratio of distinguished parcels versus the number of information points under different resolutions.

**5.2.3. Impact of communication radius.**

Figure 9 shows the comparison of the monitoring efficiency versus the number of nodes, including the parcel nodes and relay nodes, under different communication radii. We observe that large communication radius brings high monitoring efficiency and thereby the few nodes. However, large communication radius implies the bad channel quality and large energy expenditure of the relay nodes. Therefore, the proper communication radius is necessary to be derived for maintaining the high quality of data transmission and low energy consumption of the relay nodes.



**Figure 9.** Comparison of monitoring efficiency versus the number of parcel nodes under different communication radius.

**5.3. Validation of our proposed method**

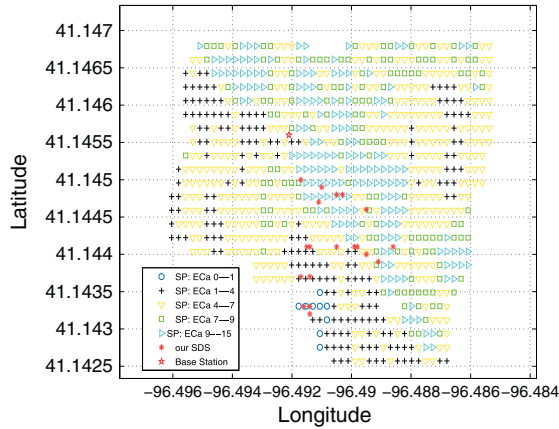
Considering the field information of water storage capacity and irrigation requirements of the field, the deviation  $\delta$  of information points of all the parcels are restricted by 0.5. The field is partitioned with different resolutions as  $a = 10, 15, 20, 25,$  and  $30$  m, respectively. Table II shows the results of the mean standard deviation of information points obtained with Equations(5) and (6). In this

**Table II.** The deviations of electrical conductivity values under different parcel sizes.

Parcel size: $a$ (m)	Deviation $d$
10	0.4337
15	0.5273
20	0.6375
25	0.7365
30	0.7986

table, the mean standard deviation when  $a = 10$  m is  $\lambda = 0.43 < \delta = 0.5$ , whereas the standard deviations of  $a = 15, 20, 25,$  and  $30$  m are all larger than  $\delta = 0.5$ . Therefore, the field is partitioned with  $a = 10$  m. To distinguish the parcels, on the basis of the extensive experiments, the range [0.23, 14.56] of the ECa values of this field is discretized into 150 information points, that is,  $y_0 = 0.23 < y_1 < y_2 < \dots < y_{150} = 14.56$  with step size  $\tau = 0.096$ , because this step size can sufficiently





**Figure 10.** The distribution of our obtained sensor deployment scheme in the field, where the electrical conductivity values are simply represented by five intervals with different colors, and each color zone has several parcel nodes deployed in it.

distinguish the most of the parcels. Table I shows the maximum achieved communication radius of 272 m at height 5. However, in Figure 9, it is worth noting that the monitoring efficiency at  $R = 80$  m is almost identical as that at  $R = 272$  m. Because short communication radius implies high quality of data transmission and long distance data transmission requires more harvested energy for sensor nodes,  $R = 80$  m is selected for data transmission among the deployed nodes. Algorithm 1 is applied to derive the approximate sensor deployment scheme, shown as the red stars in Figure 10, where the sink node is deployed in a given position.

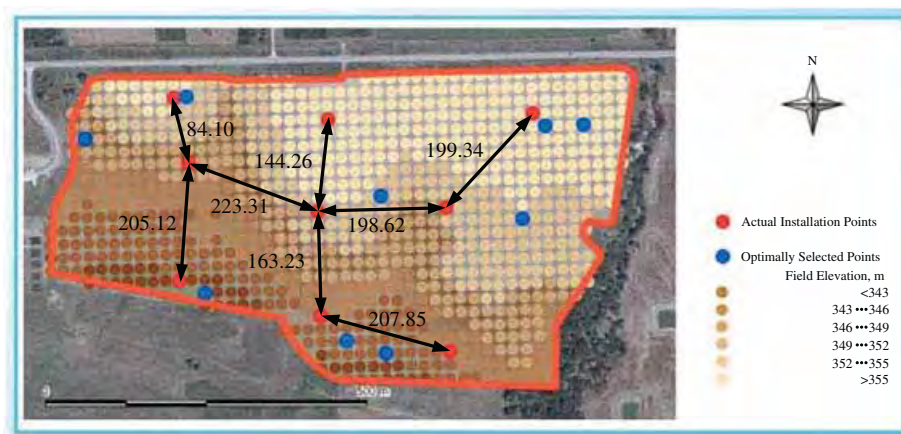
Further, we evaluate the performance of our proposed algorithms by comparing with our previous method [14]. Here, we carry out two sets of experiments. First, considering that there are nine nodes in our previous sensor deployment scheme, we also search for a nine-node sensor deployment scheme by using our proposed method.

Our nine-node sensor deployment scheme achieves the monitoring efficiency ( $\eta$ ) of 81.66% with 6.76% performance gain over our previous sensor deployment scheme. Second, because monitoring efficiency  $\eta$  of our previous scheme is 74.90% in the monitoring efficiency  $\eta$ , we investigate the minimum number of nodes that can reach the same or even better performance in regard to the same monitoring efficiency 74.90%. By utilizing the proposed method, our seven-node sensor deployment scheme reaches 70.88% in the monitoring efficiency, whereas our eight-node sensor deployment scheme reaches 75.43%. Therefore, our proposed method achieves better performance in monitoring efficiency with fewer nodes than our previous method.

As shown in Figure 11, under communication radius  $R = 80$  m, our previous nine-node sensor deployment scheme requires deploying additional 12 relay nodes to relay the collected data to the sink node. Then, the total number of nodes is 21. With our proposed method, we show the results of the performance in monitoring efficiency under different number of nodes from 10 to 21, as shown in Table III. In this table, our proposed method can meet different requirements on monitoring efficiency. Therefore, our proposed method achieves significant improvement to the previous method and thereby is more flexible to various applications.

**Table III.** The monitoring efficiency  $\eta$  and the gains of our sensor deployment schemes with different node numbers.

Node number	Monitoring efficiency ( $\eta$ )	Gain (%)
10	0.8384	8.94
11	0.8541	10.51
13	0.8773	12.83
15	0.9279	17.89
17	0.9405	19.15
19	0.9881	23.91
21	0.9997	25.07



**Figure 11.** The distances among parcel nodes in our previous sensor deployment scheme.

## 6. CONCLUSIONS

In this paper, we have proposed a novel notion of information coverage. On the basis of this notion, we have partitioned the field into multiple parcels and selected some parcels for deploying sensor nodes such that the covered field information meets the requirement. Without considering communication, we have first developed two effective polynomial-time algorithms to determine the deployed locations of parcel nodes for the information 1-coverage and  $q$ -coverage, respectively. We have also provided the theoretical bound for our proposed algorithms. Then, a polynomial-time algorithm has been developed for determining the deployed locations of relay nodes, which achieves the minimized number of relay nodes. Finally, we have successfully applied the proposed information 1-coverage in the project of WSN-based irrigation system in precision agriculture. Extensive experimental results show the major trade-offs of impact factors in sensor deployment and the achieved significant improvement of our proposed method. In the future, we will carry out the research on developing a more efficient method for processing the field information instead of information discretization and improving the approximate information 1-coverage or  $q$ -coverage solution derived by the proposed scheme in this paper.

## APPENDIX

*Proof of Proposition 1.* Suppose the range of information points is divided into  $\alpha$  intervals, where  $\alpha < \log_2 n - 1$ . As shown in Figure 2, vector  $\phi_j$  for parcel  $\mathcal{L}_j$  has  $\alpha + 1$  elements. Because each information point  $y_i$  has two possible values 0 and 1 at position  $i$  of  $\phi_j$ , we totally have  $2^{\alpha+1}$  combinations of information points. Then because each combination corresponds to a vector, we totally have  $2^{\alpha+1}$  different vectors. Considering that  $\alpha < \log_2 n - 1$  and  $2^{\alpha+1} < 2^{\log_2 n} = n$ , two identical vectors must exist among all  $2^{\alpha+1}$  vectors. This contradicts with the assumption that any two vectors are different. Therefore, the proposition holds.  $\square$

*Proof of Proposition 2.* Without losing generality, let  $\pi_j^- < \pi_k^-$  and  $\pi_j^+ < \pi_k^+$ . If at least one information point falls in  $[\pi_j^-, \pi_k^-]$  or  $[\pi_j^+, \pi_k^+]$ , the step size  $\tau$  should satisfy  $\tau \leq \max\{|\pi_k^+ - \pi_j^+|, |\pi_k^- - \pi_j^-|\}$ . Therefore, the proposition holds.  $\square$

*Proof of Proposition 5.* Let  $r = |\text{OPT}|$ , and let  $I_t$  be the set of information points of  $\{y_i\}_{i=1}^m$ , which are not yet included after step  $t$ , with  $I_0 = \{y_i\}_{i=0}^m$ . The OPT covers every  $I_t$  with no more than  $k$  parcels. Algorithm 1 always selects the parcel containing the largest number of uncovered information points over  $I_t$  in step  $t + 1$ . Hence, the size of this parcel must cover at least  $|I_t|/r$  information

points in  $I_t$ , where  $|I_t|$  represents the number of information points in  $I_t$ . If this largest parcel covers fewer points, it is impossible to cover  $I_t$  with  $r$  parcels, which contradicts the existence of OPT. So, we have  $|I_{t+1}| \leq |I_t| - |I_t|/r$ , and, inductively,

$$\begin{aligned} |I_t| &\leq |I_{t-1}| - |I_{t-1}|/r = |I_{t-1}|(1 - 1/r) \\ &\leq |I_{t-2}|(1 - 1/r)^2 \\ &\dots\dots \\ &\leq |I_0|(1 - 1/r)^t \\ &= m(1 - 1/r)^t \end{aligned} \quad (16)$$

When  $|I_t| < r = |\text{OPT}|$ , we solve for this  $t$  as follows:

$$m \left(1 - \frac{1}{k}\right)^t = r \quad (17)$$

Because  $(1 - x)^{1/x} \leq \frac{1}{e}$  for all  $x$ , we have

$$t \leq r \ln \frac{m}{r} \quad (18)$$

Thus, after  $r \ln \frac{m}{r}$  steps, there remain only  $r$  information points. Because each subsequent step removes at least one point, so  $|\text{ALG}| \leq |\text{OPT}|(\ln \frac{m}{|\text{OPT}|} + 1)$ , where ALG is the approximate sensor deployment scheme found by Algorithm 1.  $\square$

*Proof of Proposition 6.* We envision each connected component as a single node, denoted as  $\mathcal{C}_j$  by  $v_j$ , and any of the two nodes  $v_j$  and  $v_k$  has an edge  $e_{jk}$ . The cost associated with the edge  $e_{jk}$  is denoted as  $c_{jk}$ . Then, we derive a connected graph  $G$ , and the objective of Algorithm 3 is to find the minimum spanning tree  $\mathbb{Q}$ .

Suppose that Algorithm 3 produces a tree  $\mathbb{Q}$ . Let  $\mathbb{P}$  be another spanning tree with a smaller total cost. Let  $e$  be an edge of smallest weight, which lies in  $\mathbb{Q}$  but not  $\mathbb{P}$ . If we add  $e$  to  $\mathbb{P}$ , we obtain a cycle, from Equivalent Definitions for Tree. This cycle contains an edge  $e'$ , which is in  $\mathbb{P}$  but not  $\mathbb{Q}$ ; otherwise,  $\mathbb{Q}$  would not be a tree. So, we replace  $e'$  in  $\mathbb{P}$  with  $e$  from  $\mathbb{Q}$  and obtain a new spanning tree  $S'$ . From the method of construction of  $\mathbb{Q}$ , it follows that the cost of  $e$  can not exceed that of  $e'$ . So the total weight of  $S'$  does not exceed the total weight of  $\mathbb{Q}$ . Also,  $S'$  has one more edge in common with  $\mathbb{Q}$  than  $\mathbb{P}$  has.

We repeat the aforementioned procedure, and repeatedly change edges of  $\mathbb{P}$  for edges of  $\mathbb{Q}$ , and each time the weight of the intermediate graph does not exceed that of  $\mathbb{Q}$ . Then, the cost of  $\mathbb{Q}$  does not exceed that of  $\mathbb{P}$ , contradicting the definition of  $\mathbb{P}$ . Therefore,  $\mathbb{Q}$  must be a minimum spanning tree; that is, scheme  $\mathcal{S}$  derived from Algorithm 3 achieves the least number of relay nodes for connecting the connected components.  $\square$

## ACKNOWLEDGEMENTS

This work was supported in part by funds provided through Nebraska Centre for Energy Science Research, through the Water, Energy and Agriculture Initiative (WEAI), by the National Science Foundation under grant nos. 1145596 and 0830493 and the National Natural Science Foundation of China under grant no. 61302031.

## REFERENCES

1. Wang B, Chua KC, Srinivasan V, Wang W. Information coverage in randomly deployed wireless sensor networks. *IEEE Transactions on Wireless Communications* 2007; **6**(8): 2994–3004.
2. Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E. Wireless sensor networks: a survey. *Computer Networks* 2002; **38**(4): 393–422.
3. Xu X, Sahni S. Approximation algorithms for sensor deployment. *IEEE Transactions on Computers* 2007; **56**: 1681–1695.
4. Xu X, Sahni S, Rao N. Minimum cost sensor coverage of planar regions, In *Proceedings of 11th International Conference on Information Fusion*, Cologne, Germany, 2008; 1–8.
5. Yang G, Qiao D. Multi-round sensor deployment for guaranteed barrier coverage, In *Proceedings of IEEE International Conference on Computer Communications*, San Diego, USA, 2010; 1–9.
6. Shu H, Liang Q. Fuzzy optimization for distributed sensor deployment, In *Proceedings of IEEE Wireless Communications and Networking Conference*, New Orleans USA, 2005; 1903–1908.
7. Wu C-H, Lee K-C, Chung Y-C. A Delaunay triangulation based method for wireless sensor network deployment, In *Proceedings of 12th International Conference on Parallel and Distributed Systems*, Minneapolis, USA, 2006; 1–8.
8. Li S, Xu C, Pan W, Pan Y. Sensor deployment optimization for detecting maneuvering targets, In *Proceedings of 8th International Conference on Information Fusion*, Philadelphia, USA, 2005; 1629–1635.
9. Chakrabarty K, Iyengar SS, Qi H, Cho E. Grid coverage for surveillance and target location in distributed sensor networks. *IEEE Transactions on Computers* 2002; **51**(12): 1448–1453.
10. Wang G, Cao G, Porta TFL. Movement-assisted sensor deployment. *IEEE Transactions on Mobile Computing* 2006; **5**: 640–652.
11. Xu X, Sahni S. Approximation algorithms for sensor deployment. *IEEE Transactions on Computers* 2007; **56**(12): 1681–1695.
12. Zhang Y, Wang L. A sensor deployment algorithm for mobile wireless sensor networks, In *Proceedings of Chinese Control and Decision Conference*, Guilin China, 2009; 4606–4611.
13. Ma M, Yang Y. Adaptive triangular deployment algorithm for unattended mobile sensor networks. *IEEE Transactions on Computers* 2007; **56**: 946–847.
14. Adamchuk VI, Pan L, Marx DB, Martin DL. Locating soil monitoring sites using spatial analysis of multi-layer data, In *Proceedings of World Congress on Social Simulation*, Brisbane, Australia, August 2010; 1–6.
15. Xiao L, Guo L. The realization of precision agriculture monitoring system based on wireless sensor network, In *Proceedings of International Conference on Computer and Communication Technologies in Agriculture Engineering*, Chendu China, 2010; 89–92.
16. Wu B, Meng J, Zhang F, Du X, Zhang M, Chen X. Applying remote sensing in precision farming—a case study in Yucheng, In *Proceedings of World Automation Congress*, Kobe, Japan, 2010; 1–6.
17. Zhang N, Wang M, Wang N. Precision agriculture—a worldwide overview. *Computers and Electronics in Agriculture* 2002; **36**(2): 113–132.
18. Sadler EJ, Evans RG, Stone KC, Camp CR. Opportunities for conservation with precision irrigation. *Journal of Soil and Water Conservation* 2005; **60**: 371–379.
19. An W, Shao F-M, Meng H. The coverage-control optimization in sensor network subject to sensing area. *Computers and Mathematics with Applications* 2009; **57**: 529–539.
20. Ganjegunte G, Braun R. Application of electromagnetic induction technique for soil salinity and sodicity appraisal, In *Proceedings of International Conference on Environmental Engineering and Applications*, Singapore, 2010; 277–280.
21. Corwin DL, Lesch SM. Apparent soil electrical conductivity measurements in agriculture. *Computers and Electronics in Agriculture* 2005; **46**: 11–43.
22. Farahani H, Buchleiter GW, Brodahl MK. Characterization of apparent soil electrical conductivity variability in irrigated sandy and non-saline fields in Colorado. *Transactions of the ASAE* 2004; **48**(1): 155–168.
23. Shih Y, Chung W, Hsiu P, Pang A. A mobility-aware node deployment and tree construction framework for ZigBee wireless networks. *IEEE Transactions on Vehicular Technology* 2013; **62**(6): 2763–2779.
24. Wang X, Han S, Wu Y, Wang X. Coverage and energy consumption control in mobile heterogeneous wireless sensor networks. *IEEE Transactions on Automatic Control* 2013; **58**(4): 975–988.
25. Mahboubi H, Habibi J, Aghdam AG, Sayrafian-Pour K. Distributed deployment strategies for improved coverage in a network of mobile sensors with

- prioritized sensing field. *IEEE Transactions on Industrial Informatics* 2013; **9**(1): 451–461.
26. Kong L, Zhao M, Liu X, Lu J, Liu Y, Wu M, Shu W. Surface coverage in sensor networks. *IEEE Transactions on Parallel and Distributed Systems (Early View)* 2013.
  27. Martinelli M, Ioriatti L, Viani F, Benedetti M, Massa A. A WSN-based solution for precision farm purposes, In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, Cape Town, South Africa, 2009; V469–V472.
  28. Rajendran V, Obraczka K, Garcia-Luna-Aceves JJ. Energy-efficient, collision-free medium access control for wireless sensor networks, In *Proceedings of ACM Conference on Embedded Networked Sensor Systems*, Los Angeles, California, 2003; 181–192.
  29. Karp RM, Miller RE, Thatcher JW. Reducibility among combinatorial problems. *Journal of Symbolic Logic* 1975; **40**(4): 618–619.
  30. Vazirani VV. *Approximation Algorithms*. Springer-Verlag: New York, Inc.: New York, NY, 2001.
  31. Trevisan L. *Combinatorial Optimization: Exact and Approximate Algorithms*. Stanford University: San Francisco, CA, 2011.
  32. Abrams Z, Goel A, Plotkin S. Set K-cover algorithms for energy efficient monitoring in wireless sensor networks, In *Proceedings IEEE/ACM International Symposium on Information Processing in Sensor Networks*, Berkeley, California, USA, 2004; 1–9.
  33. Slijepcevic S, Potkonjak M. Power efficient organization of wireless sensor networks, In *Proceedings of IEEE International Conference on Communications*, Helsinki, Finland, 2001; 1–5.
  34. Papadimitriou CH, Steiglitz K. *Combinatorial Optimization: Algorithms and Complexity*. Dover Publications, Inc.: Mineola, NY, 1998.
  35. eKo PRO Seris-eKo Node for Environment Monitoring. Available from: <https://www.memsic.com> [accessed on 20 December 2009].

## AUTHORS' BIOGRAPHIES



**Wei An** received his BS degree in Mathematics from Linyi Normal University, Shandong, China, in 2005 and MSc and PhD degrees in Applied Mathematics and Control Science and Engineering from East China University of Science and Engineering, Shanghai, China, in 2008 and 2012, respectively. He did the research as a visiting scholar in the University of Nebraska-Lincoln, USA, from 2009 to

2011. Currently, he is a Postdoctoral Research Associate with the High Performance Network Lab of the Institute of Acoustics of the Chinese Academy of Sciences. His research interests include wireless sensor network and complex systems.



**Song Ci** [S'98-M'02-SM'06] received his BS degree from Shandong University of Technology (now Shandong University), Jinan, China, in 1992; M.S. degree from Chinese Academy of Sciences, Beijing, China, in 1998; and Ph.D. degree from the University of Nebraska-Lincoln in 2002, all in Electrical Engineering. Currently, he is an Associate Professor of Computer and Electronics Engineering at the University of Nebraska-Lincoln. His research interests include dynamic complex system modeling and optimization, green computing and power management, dynamically reconfigurable embedded system, content-aware quality-driven cross-layer optimized multimedia over wireless, cognitive network management and service-oriented architecture, and cyber-enable e-healthcare.



**Haiyan Luo** received his PhD degree in Computer Engineering from the University of Nebraska-Lincoln, USA, in 2011. In the past few years, he successfully completed several research projects, including moving object segmentation, cross-layer optimized multimedia communications over wireless networks, and wireless P2P video streaming. His current research interests include large-scale system, distributed networking technologies, service-oriented architecture, quality-driven cross-layer optimized multimedia transmission, software engineering, and next-generation network security. He is an IEEE senior member. Dr. Luo served as a TPC member in a few well-known international conferences such as IEEE GLOBECOM, IEEE ICC, IEEE WCNC, and IWCMC and a technical referee of many major technical journals such as IEEE JSAC, IEEE TWC, IEEE TMM, IEEE TNSM, and KSII Transactions on Internet and Information Systems. He is also an editor of a few internationally well-known journals. Dr. Luo is currently associated with Cisco Systems on the design and development of next generation network security platform.



**Dalei Wu** [S'05-M'10] received his B.S. and M.E. degrees in Electrical Engineering from Shandong University, Jinan, China, in 2001 and 2004, respectively, and Ph.D. degree in Computer Engineering from the University of Nebraska-Lincoln in 2010. Currently, he is a Postdoctoral Research



Associate with the Mechatronics Research Lab, Massachusetts Institute of Technology. His research interests include cyber-physical systems, wireless sensor networks, and multimedia computing and networking.



**Viacheslav Adamchuk** obtained a Mechanical Engineering degree from the National Agricultural University of Ukraine in 1996. He has received both M.S. (1998) and Ph.D. (2000) degrees in Agricultural and Biological Engineering from Purdue University. Soon after graduation, he started his academic career as a faculty member in Biological Systems Engineering Department at the University of Nebraska-Lincoln. Now, he is an Associate Professor in Bioresource Engineering Department at McGill University. His primary research interest is in design and analysis of automated systems for mapping soil properties on-the-go.

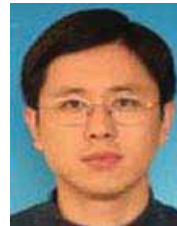


**Hamid Sharif** is the Cheryl Prewett Diamond Professor of the Computer and Electronics Engineering Department and the Director of the Advanced Telecommunications Engineering Laboratory (TEL) at the Peter Kiewit Institute in the University of Nebraska-Lincoln (Omaha Campus). He joined the University of Nebraska-Lincoln in 1986 and holds a B.S. degree from the University of Iowa, an M.S. degree from the University of Missouri-Columbia, and a Ph.D. degree from the University of Nebraska-Lincoln, all in Electrical Engineering. Dr. Sharif's research areas include network communications protocols, wireless and wireless sensor network design, the next generation of internet architecture, and network modeling and performance analysis. His research works have been supported

by national research agencies such as the National Science Foundation, Department of Defense, and Department of Energy, as well as many private organizations. He has authored or co-authored over 160 articles in refereed journals and conferences. He is a senior member of many professional organizations and the current chair of the IEEE Nebraska Section and past president of the IEEE Nebraska Computer and Communications Societies. He has served as a technical program committee member for many national/international conferences.

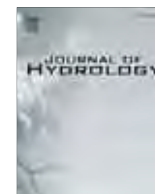


**Xueyi Wang** received his B.S. degree in Automation from Southeast University, Nanjing, China, in 2005. He received his M.S. degree in Communication and Information Systems from China University of Mining and Technology, Beijing, China, in 2008 and his MS degree in Telecommunication from University of Nebraska-Lincoln, USA, in 2011, respectively. He is currently working for FEV, Inc. on engine control and calibration. His research interests include software and hardware design for Embedded Systems, HIL System, and Engine Control Strategy.



**Dr. Hui Tang** received his B.S. degree from Lanzhou University in 1992, his M.S. degree from the Institute of Computing Technology of the Chinese Academy of Sciences in 1995, and his Ph.D. degree from the Institute of Acoustics of the Chinese Academy of Sciences in 1998. Since 2004, he has become the founding director of the High Performance Network Laboratory of the Institute of Acoustics of the Chinese Academy of Sciences. His research interest is the next-generation broadband wireless mobile network.





## Assessing the impacts of the urban heat island effect on streamflow patterns in Ottawa, Canada



Jan Adamowski<sup>a,\*</sup>, Andreas Prokoph<sup>b</sup>

<sup>a</sup> Department of Bioresource Engineering, McGill University, 21 111 Lakeshore Road, Ste. Anne de Bellevue, QC, H9X 3V9 Canada

<sup>b</sup> SPEEDSTAT, 19 Langstrom Crescent, Ottawa, ON, K1G 5J5 Canada

### ARTICLE INFO

#### Article history:

Received 10 March 2013

Received in revised form 13 May 2013

Accepted 15 May 2013

Available online 23 May 2013

This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Ashish Sharma, Associate Editor

#### Keywords:

Wavelet transform

Urban heat island effect

Climate change

Urban water

Streamflow

Canada

### SUMMARY

Due to a variety of commercial and residential activities, large metropolitan areas in mid-to-high latitudinal ranges are experiencing rising air temperatures compared to their surrounding rural areas. This study investigated how this urban heat island effect (UHIE) may influence the streamflow of rivers crossing large urban areas on annual and multi-decadal time-scales. In order to detect, link, and quantify differences in meteorological and streamflow patterns between rural and large urban areas, this study developed a methodology based on the continuous wavelet transform (CWT), cross-wavelet transform (XWT), linear regression, as well as the Mann–Kendall (MK) test. A case study was carried out for the city of Ottawa, Canada as the metropolitan centre, along with three surrounding rural locations (Angers, Arnprior, Russell), with pristine rivers crossing these locations. From roughly 1970 to 2000, air temperature in Ottawa increased at a rate exceeding 0.035 °C/year, while parallel changes in rural areas were relatively stable, and varied by less than 0.025 °C/year. The urban warming that occurred during these decades was accompanied by a significant drop in the amplitude of annual temperatures (i.e. warmer winters). Precipitation in both urban and rural areas showed no significant trends, although the variability in the precipitation amount decreased in both settings. Concurrently, streamflow showed decreasing trends in both urban and rural areas. Annual amplitudes in urban streamflow (Rideau River through Ottawa, ON) correlated positively with annual air temperature amplitudes (i.e., less severe annual flooding with a decreasing winter/summer temperature contrast), whereas such a relationship was not apparent for the rural stations. Moreover, the timing of the annual daily minimum temperature cycle correlated significantly with the streamflow pattern in the urban area, i.e., early annual warming corresponded to earlier annual streamflow maxima. The precipitation pattern (i.e. distribution of rain and snowfall over time) significantly influenced the annual and long term streamflow pattern, but this influence differed little between urban and rural areas. It was also determined that the warming from the urban heat island effect, especially during winter months, was found to perhaps reduce the severity of the annual spring flood event in mid-to-high latitudinal continental settings.

© 2013 Elsevier B.V. All rights reserved.

### 1. Introduction

The expansion of urban centers has created, over the years, a pronounced warming in urban areas relative to their rural surroundings (Oke, 1973; Karaca et al., 1995; Arnfield, 2003). There are several mechanisms caused by urbanization that can influence streamflow. This includes human induced changes such as changes through construction, river flow regulation to water outfall and heated water disposal, as well as natural factors that are influenced by urbanization. These mechanisms include altered natural causes such as ambient air temperature increases and altered precipitation

patterns in urban centers. These altered natural mechanisms can lead to earlier snowmelt in the year, as well as changes in precipitation related runoff resulting in changes in temporal streamflow patterns compared to the rural surroundings. These differences in air temperatures in city centers, first observed by Howard (1833) as early as the 19th century, were eventually coined under the term “Urban Heat Island” (UHI) by Manley (1958), when he was investigating changes in snowfall patterns between rural and urban areas. Over the years, research has shown that as population increased and cities grew, so did the intensity of their UHIE (Oke, 1973; Li et al., 2004). It has also been observed that the UHIE provides additional warming to both an urban center and its immediate surrounding areas, and that this warming is more pronounced during the winter months when commercial and residential heating are at their highest (Karl et al., 1988).

\* Corresponding author. Tel.: +1 5143987786.

E-mail addresses: [jan.adamowski@mcgill.ca](mailto:jan.adamowski@mcgill.ca) (J. Adamowski), [aprokocon@aol.com](mailto:aprokocon@aol.com) (A. Prokoph).

UHIs have been studied in various parts of the world (Arnfield, 2003), with locations ranging as far as Alaska in northern latitudes (Hinkel et al., 2003) and in warmer and drier regions such as Spain and Turkey (e.g., Yague and Zurita, 1991; Karaca et al., 1995; Montáñez et al., 2000; Yalcin and Yetemen, 2009). With regards to Canada, Oke has investigated temperature trends in urban centers in the St-Lawrence Lowlands as well as in the Pacific Northwest (Oke, 1973; Oke and Maxwell, 1975), with current research now focused in detecting correlations between meteorological conditions and the intensity of the UHI phenomenon in central Canadian cities such as Toronto and Regina (Stewart, 2000; Moshin and Gough, 2012).

While research on the UHIE has been previously geared towards identifying the effect of urban warming on ambient air temperatures, there has been in recent years some research that investigates the possible effect of urban heating on water resources. For example, Shepherd and Burian (2003) and Lin et al. (2011) have examined the impact of UHIs on rainfall anomalies in coastal areas. Their research showed that urban centers located near water masses are becoming warmer and drier because higher rates of evaporation in urbanized areas are blocking precipitation formation by inhibiting water vapor from being transported from the coast. Kinouchi et al. (2007) also investigated the effect of UHIs on water sources by looking at increases in stream temperatures related to heat inputs from urban wastewater. Their study revealed that there was a correlation between increases in urban stream temperatures and increases in urban wastewater temperatures, which they expect to increase even more in the future due to higher urbanization rates and a rise in energy demand and water consumption. More recently, Yalcin and Yetemen (2009) have brought forward the idea that UHIs can also have an impact on groundwater resources. From their analysis of temperature data in underground layers of streams and wells near Istanbul, they observed that the water temperatures in urban groundwater sources were on average 3.5 °C higher than the rural groundwater sources.

Whereas it can be seen from the above examples that research involving the UHIE on water resources is slowly gaining momentum, there is still a lack of studies that directly examine the effect of UHIs on urban streamflow patterns and variability, with past research having been predominately focused on rural streamflows (Lettenmaier et al., 1994; Krakauer and Fung, 2008). Zhang et al. (2001) determined that, in general, streamflow in Canadian rivers had decreased over the last 30–50 years, except for the months of March–April when spring streamflow had increased. Projected global climate warming over the coming century will clearly influence streamflow patterns in snowmelt-induced annual discharge cycles typical of mid-to-high-latitude continental climate settings, especially with regards to the patterns' magnitude and onset (Douglas et al., 2000). Many large, heavily populated urban centers such as Ottawa, Minneapolis, Chicago, Berlin, Kiev and Moscow are located in mid to high latitude continental settings. It thus becomes very important to characterize the effect of UHIE on the streams that are found within these particular urban centers.

Both wavelet analysis and the Mann–Kendall test have been used (although not together) to determine streamflow trends and other patterns, as well as to analyze climate records in Canada. The Mann–Kendall test (Yulianti and Burn, 1998) and wavelet analysis (e.g., Anctil and Coulibaly, 2004; Abdul Aziz and Burn, 2006; Burn et al., 2008; Adamowski et al., 2009) have been predominately used in determining long-term (multiannual) trends and relationships. These studies found a positive correlation between streamflows and patterns of both the El-Nino/Southern Oscillation (ENSO) and the North Atlantic Oscillation (NAO). They also revealed a shift in these streamflow patterns between 1950 and 1970. Nakken (1999) also used wavelet analysis to investigate the potential anthropogenic influences on streamflow, focusing

specifically on the rainfall-streamflow pattern, and finding a strong relationship between them.

The purpose of this study is twofold: (i) Develop a criterion to determine and quantify the influence of overlying long-term regional and global climate trends on river flow patterns. Waveband-specific trends and climate-streamflow relationships which have been found to occur can be best addressed by using wavelet and cross-wavelet transform-based amplitude and phase-lag extraction techniques and significance of trend evaluations using Mann–Kendall tests and assessments of the significance of linear correlations. The aim is to determine if streamflow is significantly influenced by UHIE and, if so, at which wavelengths (time-scale). (ii) Complete a case study for Ottawa, Canada focusing on multi-decadal vs. annual timescale impacts of meteorological variations on river flow patterns, based on continuous daily records from 1972 to 1998. Drawing on several rural and urban monthly climate records, Prokoph and Patterson (2004) used wavelet and trend analysis to determine the impact of the UHIE in the Ottawa area. Over the last fifty years, UHIE warming in Ottawa has occurred at a mean rate of 0.009 °C/yr, and has, in turn, strongly raised regional background warming to a rate of roughly 0.006 °C/yr. Particularly large increases in urban–rural temperature differences occurred during periods of accelerated population growth in Ottawa. Their results provide the basis for the climate record analysis undertaken in this study.

## 2. Methodological approach

The ability to access extensive streamflow and climate records (e.g. temperature, total precipitation) for Ottawa and its surrounding areas was an important aspect of this study. The records provided information on daily streamflow changes for rivers flowing from a rural area into a high heat-island-affected city centre. For this study, one pair of daily and monthly records from urban streamflow/meteorological stations, and, as a control set, a rural streamflow/meteorological pair of station records, were chosen for detailed comparison. Data from Environment Canada (weather.ec.gc.ca, and ec.gc.ca/rhc-wsc) was used as the primary source of data for these observations. In this case study, the streamflow discharge ( $F$ ), both at daily- and monthly-averaged sampling rates were used as hydrological dependent records, while different types of precipitation ( $P$ ) and air temperature records served as independent records. Other climate-related records such as evapotranspiration, snow-on ground/day, snowmelt/day or daily and monthly extreme values were not available at sufficient continuity to be used for this study. Monthly mean temperature ( $T$ ), daily maximum temperature ( $T_{\max}$ ) and daily minimum temperature ( $T_{\min}$ ) were used in this study. Monthly records are sufficient to analyse the general annual pattern, but important information is lost concerning diurnal fluctuations, represented by  $T_{\min}$  and  $T_{\max}$ , that are important in characterizing UHIE (e.g., Karl et al., 1988). Ten methodological steps were carried out in this study to determine the potential influence of the UHIE on long-term and annual streamflow patterns (Fig. 1):

- (i) Gathering pairs of nearby (<5 km apart) streamflow and climate records over the same time-interval, at the same sampling rate, and preferably with perfect completeness. These records were then grouped into potential UHIE influenced (urban) and non-influenced (rural) records.
- (ii) Calculating descriptive statistics for each record to provide an overview of distribution and extreme values. This process was aided by visual observation of the plotted records.
- (iii) Calculating and evaluating each record's sample distribution with regard to Gaussianity, a prerequisite for optimal signal detection and extraction using Fourier analysis and the

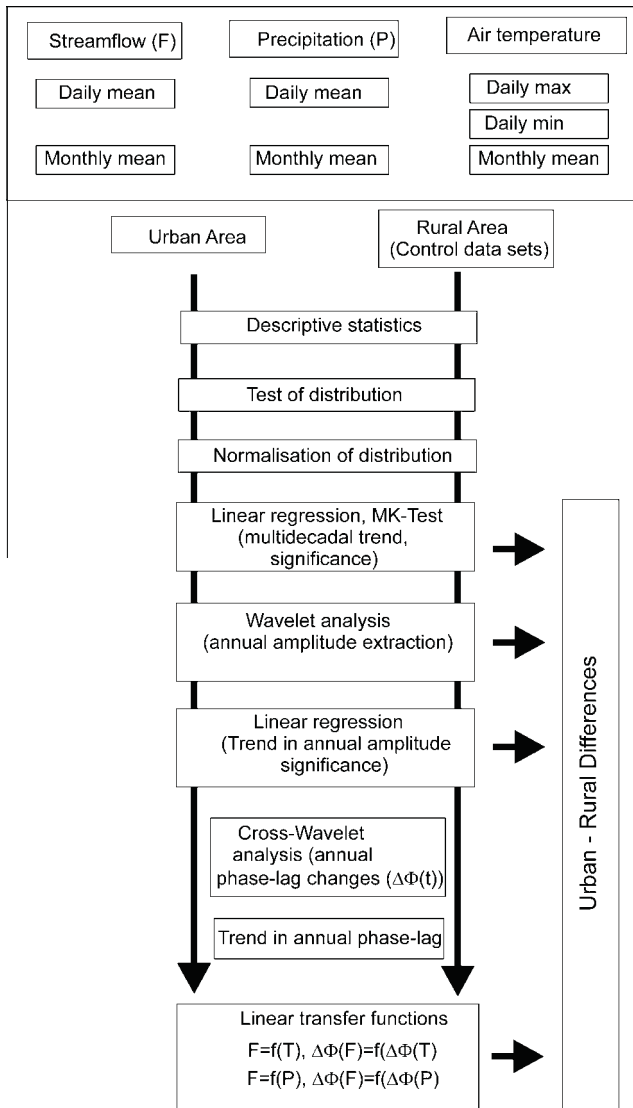


Fig. 1. Flow diagram for processing of both urban heat island affected "URBAN" and unaffected ("RURAL") areas.

wavelet transform. In this study, the latter was done using the Morlet wavelet. It should be noted that the distribution may not be scale-invariant, i.e., a record may be non-Gaussianly distributed on a short time-scale but Gaussianly distributed on a long time-scale. We tested Gaussianity using the histogram methods of Al-Smadi (2005).

- (iv) Transforming non-Gaussian distributed records into Gaussian distributed records using lognormal or other case-specific transformation techniques adequate for hydrological data (Granato, 2009).
- (v) Determining long-term, usually multi-decadal trends for less than 100-year record lengths using linear regression, then testing the significance of the linear trend, using the non-parametric Mann–Kendall Test (Kendall, 1975). Consequently, records should preferably have exactly the same observation/sampling period.
- (vi) Extracting annual signals, in particular amplitudes as a function of time, using CWT (Prokoph and Barthelmes, 1996). In contrast to spectral analysis, CWT allows for the extraction of not only single (averaged) signal amplitudes, or spectral power and phase per waveband, but also time-dependent records which allow for the tracing of the signal through different urbanization stages.

- (vii) Applying linear regression and the MK Test with related significance testing on annual amplitude records.
- (viii) Applying cross wavelet analysis (XWT) to hydrological and climate data to extract changes in the annual phase-offset between paired urban and rural records.
- (ix) Determining whether the annual urban–rural phase-relationship remained stable or not over time, and using linear regression to determine if a trend occurs in the annual phase-lag changes.
- (x) Determining: (a) The significance of long-term and annual-scale relationships between climate–streamflow, and (b) if significant, differences in trends between urban–rural records were assessed by cross-plots and Pearson correlation coefficients. The annual amplitude and phase-lag record were downsampled to one data point per year to ensure non-autocorrelated, independent samples for significance evaluation. Differences in trends determined whether the influence of the UHIE on streamflow had an impact on long-term streamflow trends, annual amplitude of streamflow, or timing (phase-offset) of annual streamflow cycle fluctuations.

### 3. Data analysis methods

#### 3.1. Continuous wavelet analysis

The Continuous Wavelet Transform (CWT) allows for the automatic localization of periodic signals, gradual shifts and abrupt interruptions, trends and onsets of trends in time series (Rioul and Vetterli, 1991). In contrast to Fourier analysis, CWT permits the transformation of observed time series into wavelet coefficients according to time and scale (or frequency) simultaneously. These coefficients can be used to detect and estimate trends or to reconstruct signals in streamflow and meteorological records that are of interest for this study. Wavelet analysis first emerged as a filtering and data compression method in the 1980s (e.g., Morlet et al., 1982a,b). Wavelet analysis transforms a time-series into a frequency domain, then simultaneously transforms the 'depth' or 'time' domain and the 'scale' or frequency domain by using various shapes and sizes of short filtering functions called 'wavelets.' The wavelet coefficients  $W$  of a time series  $x(s)$  are calculated by a simple convolution (Prokoph and Barthelmes, 1996):

$$W_{\psi}(a, b) = \left( \frac{1}{\sqrt{a}} \right) \int x(s) \psi \left( \frac{s-b}{a} \right) ds \quad (1)$$

where  $a$  is the scale factor that determines the characteristic frequency or wavelength,  $b$  represents the shift of the wavelet over  $x(s)$ , and  $\psi$  is the mother wavelet; the variable.

The bandwidth resolution for a wavelet transform varies with  $\Delta a = \frac{\sqrt{2}}{4\pi a^l}$ , and a location or time resolution  $\Delta b = \frac{a^l}{\sqrt{2}}$ . Note that due to Heisenberg's uncertainty principle  $\Delta a \Delta b \geq 1/4\pi$ , and the resolution of  $\Delta b$  and  $\Delta a$  cannot be arbitrarily small (e.g., Prokoph and Barthelmes, 1996). Parameter  $l$  is used to modify wavelet transform bandwidth resolution either in favour of time or in favour of frequency. In this study, the CWT was used with the Morlet wavelet as the mother function (Morlet et al., 1982a,b). The shifted and scaled Morlet mother wavelet is defined as:

$$\psi_{l,a,b}(s) = \sqrt[4]{\pi} \sqrt{a} e^{-\frac{i2\pi(s-b)}{a}} e^{-\frac{1}{2} \left( \frac{s-b}{a} \right)^2} \quad (2)$$

The Morlet wavelet is simply a sinusoid with wavelength/period  $a$  modulated by a Gaussian function (Torrence and Compo, 1998; Prokoph and Patterson, 2004; Adamowski, 2008). Edge effects of the wavelet coefficients occur at the beginning and end of the analysed time-series and increase with increasing wavelength (scale) and with the  $l$  parameter forming a 'cone of influence of edge ef-

fects' (about 5% in this study) (Torrence and Compo, 1998). The wavelet coefficients  $W$  are represented in this study by the amplitude of Fourier frequencies by replacing  $\sqrt{a}$  with  $a$ . The parameter  $l=6$  was chosen for all analyses, which gives sufficiently precise results in resolution of depth and frequency, respectively (Ware and Thomson, 2000; Adamowski, 2008). The wavelet analysis technique used in this article is explained in greater detail by Prokoph and Barthelmes (1996) and Adamowski et al. (2009).

In this study, wavelet coefficients were calibrated to reduce their exponential decay due to edge effects by dividing the wavelet coefficient of wavelength  $a$  extracted from Eq. (1) by a standing sine wave of amplitude 1 and wavelength of 365 days. The matrix of the wavelet coefficients  $W(a,b)$ , or the so called 'scalogram', was coded with shades of grey for superior graphical interpretation. The scale with the strongest wavelet coefficient in the 320–400 day waveband, and its associated wavelet coefficient and phase, were extracted for each time interval  $\Delta b$ .

The differences in CWT compared to spectral analysis for the purpose of this study (extractions of records of amplitudes of annual signals) are illustrated in Fig. 2. A set of three deterministic models with record lengths in model units (mu)  $t=0 \dots 200$  were constructed:

**Model 1:**  $y(t) = 0.01t[1 + \cos(\frac{2\pi t}{12})]$  implements a stationary annual signal of gradually increasing amplitude with the addition of an underlying linear trend. This model can represent long-term gradually-increasing streamflow, as well as increased annual low-high streamflow amplitudes.

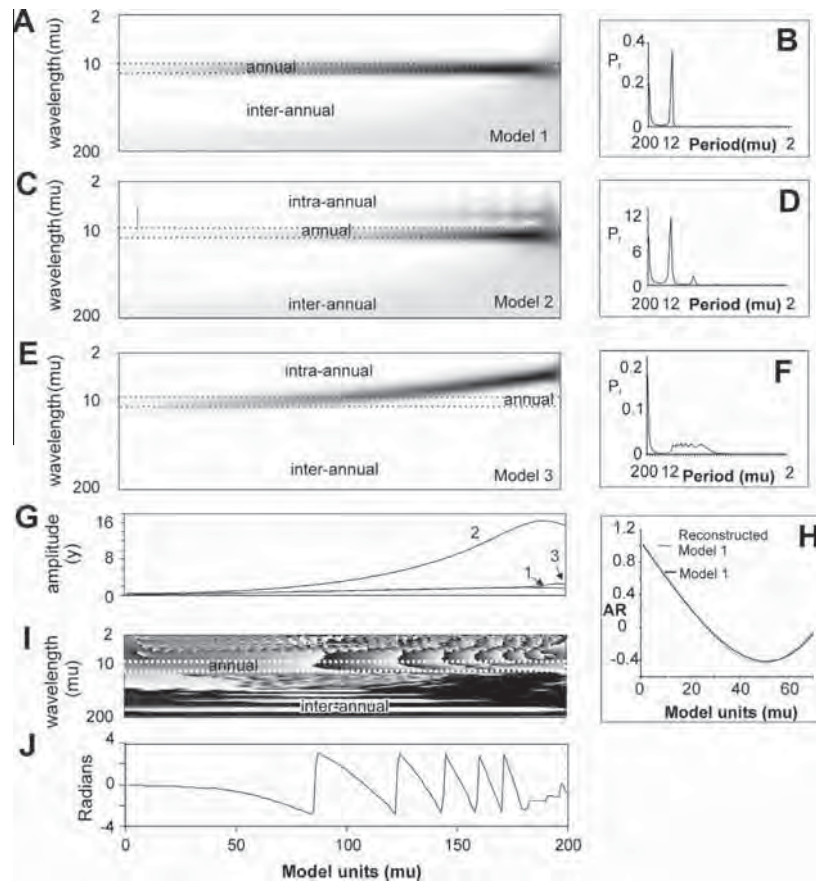
**Model 2:**  $y(t) = \exp(0.01t)[1 + \cos(\frac{2\pi t}{12})]$  has the same structure as model #1, but is exponentially enhanced.

**Model 3:**  $y(t) = 0.01t[1 + \cos(\frac{2\pi t}{12-0.01t^2})]$  represents a 'chirp' (Mann and Haykin, 1992) with a synchronous increase of amplitude and frequency due to a resonance effect, which can represent an earlier arrival of the annual snowmelt-induced flood paralleled by increasing flood amplitude.

In contrast to spectral analysis (Fig. 2B, D and F), CWT is able to trace and extract the temporal signal changes, in particular for Model 3 (Fig. 2E), where spectral analysis provides a low-spectral power distributed over a broad waveband (Fig. 2F). However, the period doubling of the annual signal in both spectral analysis and CWT with model 2 (Fig. 2C and D) illustrates the importance of transforming, in this case exponentially, a non-Gaussian distributed signal into a Gaussian-distributed record to avoid amplitude loss in the waveband of interest. Moreover, the autocorrelation of the original and the reconstructed signal is preserved (Fig. 3H). Thus, autocorrelation of amplitudes extracted from annual signals of climate or streamflow records will not be biased, except for some cases, as in model 2, and no bias will be introduced in evaluating the significance of trends.

### 3.2. Cross-wavelet analysis (XWT)

The cross-wavelet spectrum of two series  $x(t)$  and  $y(t)$  is defined by:



**Fig. 2.** Wavelet (CWT), cross wavelet (XWT) and spectral analysis (SA) to detect annual and long-term nature of UHIE for model units (mu)  $t=0 \dots 200$ . (A) CWT and (B) SA of Model 1:  $y(t) = 0.01t + 0.01t \cos(2\pi t/12)$ . (C) CWT and (D) SA of Model 2:  $y(t) = \exp(0.01t) + 0.01t \cos(2\pi t/12)$ . (E) CWT and (F) SA of a "chirp" Model 3:  $y(t) = 0.01t + 0.01t \cos(2\pi t/(12-0.01t^2))$ . (G) Extracted maximum annual amplitudes, corrected for edge effects. (H) Autocorrelations of amplitudes of Model 1 and reconstructed Model 1. (I) Phase spectrum of chirp (Model 3) versus  $y(t) = 0.01t \cos(2\pi t/12)$ . (J) Extracted phase lag of annual waveband (11–13 model units bandwidth) from (I). Wavelet scalogram using Morlet wavelet with window width of 6, striped line marks >10% edge effect influence (below line), black-dark grey: large signal, white-low/no signal with intra-annual band (5–320 days), annual band (320–400 days) and inter-annual band (400 days –  $2 \times$  total number of days).



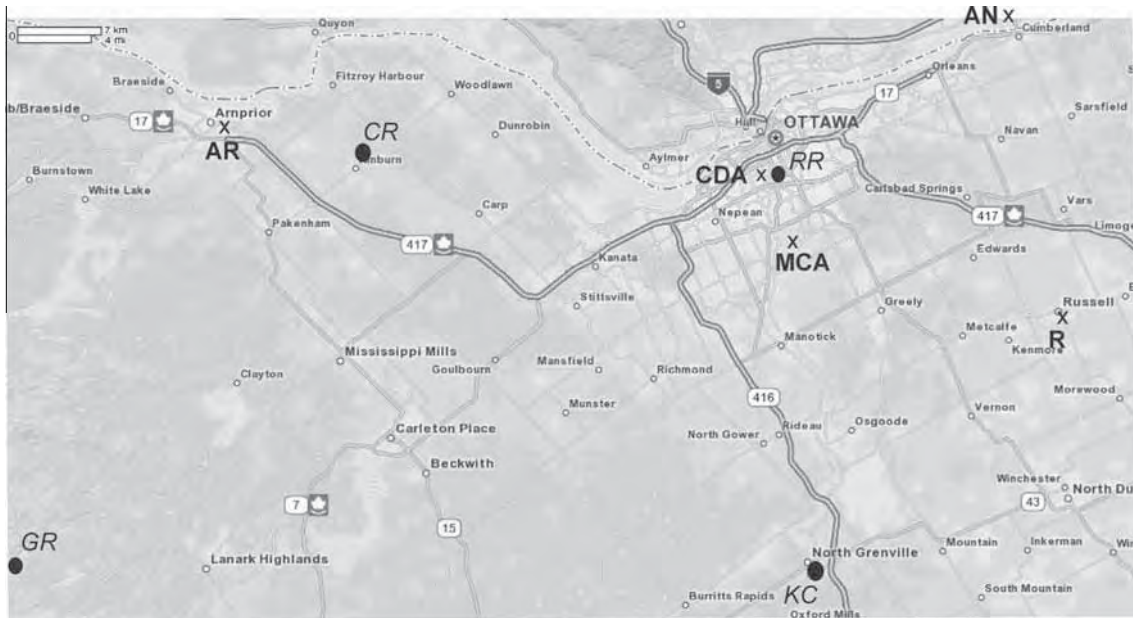


Fig. 3. Ottawa, Ontario map with meteorological stations (bold) and hydrological stations (italic): AR-Arnrior, CDA-Ottawa CDA, MCA-Ottawa MacDonald-Cartier Airport, AN-Angers, R-Russell, CR-Carp River at Kinburn, GR-Clyde River at Gordon Rapids, KC-Kemptville Creek at Kemptville, RR-Rideau River in Ottawa.

$$W_{xy}(a, b) = W_x(a, b)W_y^*(a, b) \tag{3}$$

where  $W_x(a, b)$  and  $W_y(a, b)$  are the CWT of  $x(t)$  and  $y(t)$ , respectively, and  $*$  denotes the complex conjugate (e.g., Jury et al., 2002). The phase difference is defined as:

$$\Delta\Phi(b) = \tan^{-1} \frac{\int_{a1}^{a2} \text{Im}(W_{xy}(a, b)) da}{\int_{a1}^{a2} \text{Re}(W_{xy}(a, b)) da} \tag{4}$$

where  $b$  is the time lag  $b$  (Jury et al., 2002), and  $Im$  and  $Re$  represent the imaginary and real parts of the wavelet transform, respectively. The mother wavelet and parameters used in this study are the same as for the CWT description provided above. More details on the XWT technique that was used in this study can be found in Maraun and Kurths, (2004); Labat, (2008). They found that significance testing of the cross-wavelet amplitude, or the often used cross-wavelet coherency, is ambiguous. In this study, XWT was used to extract the phase lag between two records at a specific (here, annual) waveband  $\Delta\phi(b) = \Delta\phi(t)$ . The gradual changes in the phase shift can be properly traced and extracted as demonstrated by the XWT of model 3 compared to a cosine wave with a wavelength of 12 model units (Fig. 2I and J).

### 3.3. Mann–Kendall test

The Mann–Kendall test is based on the rank correlation methodology introduced by Kendall (1975) with associated statistical tests implemented by Gilbert (1987). The Mann–Kendall test can be applied when the data values  $x_i$  of a time series can be assumed to obey the model:

$$X_i = f(t_i) + \varepsilon_i \tag{5}$$

where  $f(t_i)$  is a continuous monotonic increasing or decreasing function of time, where  $\varepsilon_i$  (the residuals), are assumed to be from the same distribution and have a mean of zero (e.g., Salmi et al., 2002). The variance in the distribution is thus assumed to be constant in time. The null hypothesis of no trend existing,  $H_0$ , (i.e., the observations  $x_i$  are randomly ordered in time) was tested against the alternative hypothesis,  $H_a$ , assuming there to be an increasing or decreasing monotonic trend. The software MAKESENS

(Salmi et al., 2002), was used to calculate the Mann–Kendall test, which is based upon the normal approximation using Z statistics.

Let  $x_1, x_2, \dots, x_n$  represent  $n$  data points where  $x_j$  represents the data point at time  $j$ . Then the Mann–Kendall statistic ( $S$ ) is given by (Kendall, 1975):

$$S = \sum_{j=1}^{j=n} \sum_{k=1}^{k=n} \text{sign}(x_j - x_k) \tag{6}$$

where  $\text{sign}(x_j - x_k)$  equals 1 if  $x_j - x_k > 0$ , equals 0 if  $x_j - x_k = 0$ , and equals  $-1$  if  $x_j - x_k < 0$

A very high positive value of  $S$  is an indicator of an increasing trend, while a very low negative value indicates a decreasing trend. However, it is necessary to compute the probability associated with  $S$  and the sample size,  $n$ , to statistically quantify the significance of the trend.

The variance (VAR) of  $S$  is computed by the following equation:

$$\text{VAR}(S) = 1/18(n(n-1)(2n+5) - \sum t_p(t_p-1)(2t_p+5)) \tag{7}$$

where  $n$  is the number of data points, and  $t_p$  is the number of data points in the  $p$ th group. A normalized test statistic  $Z$  is then calculated as follows:

$$Z = \frac{S - 1}{\sqrt{\text{VAR}(S)}} \text{ if } S > 0 \tag{8}$$

$$Z = 0 \text{ if } S = 0 \tag{9}$$

$$Z = \frac{S + 1}{\sqrt{\text{VAR}(S)}} \text{ if } S < 0 \tag{10}$$

The presence of a statistically significant trend is evaluated using the  $Z$  value. A positive value of  $Z$  indicates an upward trend, and a negative value a downward trend. The statistic  $Z$  has a normal distribution. To test for either an upward or downward monotone trend at a level of significance  $\alpha$ ,  $H_0$  is rejected if the absolute value of  $Z$  exceeds  $Z_{1-0.5\alpha}$ , which is provided by the standard normal cumulative distribution (e.g., Davis, 1986). The tested significance levels implemented in MAKESENS are  $\alpha = 0.001, 0.01, 0.05$  or  $0.1$ .



## 4. Data

### 4.1. Streamflow records

Located just south of Ottawa's centre, the Rideau River station was the primary source of streamflow data, whereas the other (non-urban) hydrological stations were located in streams in rural areas (Fig. 1). The Rideau River Authority provided digital records of two daily and four mean-monthly streamflow records; the Water Service of Canada website ([wsc.ec.gc.ca](http://wsc.ec.gc.ca)) was also used as a source of similar records. Complete records ranging from 1972–2005 were available for all stations, and some extended as far back as the 1930s (Table 1). The only streamflow records which were considered were those which showed stable land-use or pristine conditions, less than 5% missing data, and at least 20 years of properly monitored measurements, taken over the same time-interval as nearby meteorological stations. Furthermore, the records that were used are approved by the Water Service of Canada and/or the Rideau River Authority.

### 4.2. Climate records

Data was collected and accessed from various meteorological stations and weather and government sources. The three rural stations under study were located in Arnprior, Russell, and Angers, Ontario, and represented areas with rural climates. The stations were located in small towns with populations under 25,000 (Fig. 3). The urban area of Ottawa centre was represented by the Canadian Department of Agriculture (CDA) station approximately 2 km south of downtown Ottawa, and about 4 km east from the Rideau River hydrological station. The Ottawa MCA (MacDonald–Cartier Airport) station is situated in a suburban area gradually shifting from a rural to an urban setting. The meteorological data analyzed during the study focused predominately on air temperatures and precipitation levels and were provided by Environment Canada ([weather.ec.gc.ca](http://weather.ec.gc.ca)). These were daily and monthly total precipitation, monthly mean temperature, and daily minimum and maximum temperatures (Table 1, Fig. 4). Based on continuous 24 h recordings, all climate records were approved for public and academic release by Environment Canada (2009) and as such were corrected and calibrated. Estimates, below detection limit mea-

surements or missing data were recorded. Only records that were more than 95% complete (i.e., that were not estimates or missing entries) were used. Priority was given to the urban climate-streamflow pairs of Ottawa CDA–Rideau River records and the rural Arnprior–Carp River pairs based on population difference, vicinity of the climate-streamflow pair and almost equivalent recording periods.

## 5. Results

### 5.1. Descriptive statistics and sample distribution of climate and streamflow records

During the case study, differences in meteorological and streamflow patterns within Ottawa (and the three surrounding rural areas of Arnprior, Russell, and Angers) were identified and quantified. The data analysis focused on pairs of streamflow-climate records that were obtained less than 10 km apart (see Fig. 3). The urban pair was the Rideau River at Ottawa, with Ottawa CDA, and the rural pair was the Carp River at Kinburn Side Road Bridge, and Arnprior. The climate records in all locations were nearly complete from 1972 to 1999, with a total of 10 days missing in 1982 and 1985 from the Arnprior records. All streamflow records show short and high annual discharge peaks due to snow melt with a long low discharge interval in between (Fig. 4A and B). The total amount of discharge varied depending on the magnitude of streamflow, with the Rideau River discharge roughly 10 times that of the rural streamflows (Table 1). The sample distribution of all the records was very similar, as indicated by their kurtosis and skewness (Table 1). The magnitudes of monthly mean temperature fluctuations were similar between all records, but these differed significantly in terms of precipitation records (see Fig. 4C and D). Moreover, histograms indicate that the temperatures were scale invariant and normally distributed, with a slight bi-modal pattern (Fig. 5A and B). The distribution was non-Gaussian for daily precipitation records, but Gaussian for monthly records. Thus, for the analysis of annual amplitudes in this study, a Gaussian distribution can be assumed. In contrast, streamflow data were not normally distributed on either daily or monthly scales (Fig. 5E and G), confirming the non-Gaussian (e.g., log-normal) distribution in streamflow time-series found in other studies (e.g., Schaeffli and Zehe,

**Table 1**  
Hydrological and meteorological stations and descriptive statistics.

Station name	Station ID	Type	Unit	Data Range	Mean	Median	Standard deviation	Kurtosis	Skewness	Range	total data range
Rideau River at Ottawa, daily	WSC02LA004	F	m <sup>3</sup> /s	1/1/72–14/7/2005	44.25	23.70	59.40	16.15	3.44	581.52	10/12/69–31/10/09
Kemptville Creek, daily	WSC02LA006	F	m <sup>3</sup> /s	1/1/72–14/7/2005	2.89	1.89	6.36	17.51	3.66	80.09	11/12/69–28/10/09
Carp River near Kinburn, daily	WSC02KF011	F	m <sup>3</sup> /s	1/1/72–14/7/2005	5.00	0.84	8.51	36.78	5.25	85.00	4/9/71–14/7/05
Clyde River, daily	WSC02KF013	F	m <sup>3</sup> /s	1/1/72–14/7/2005	3.22	1.72	4.98	29.99	4.23	87.99	9/9/71–31/12/07
Ottawa, CDA, monthly	6105976	T	°C	1/1/72–1/12/05	6.30	7.10	11.18	−1.29	−0.21	40.30	1/1/1889–1/12/06
Ottawa, Airport, monthly	6106000	T	°C	1/1/72–1/12/05	6.18	6.95	11.14	−1.27	−0.20	40.70	1/1/1938–1/12/06
Angers, QC, monthly	7030170	T	°C	1/1/72–1/12/05	4.92	5.70	11.20	−1.26	−0.23	41.10	1/1/1962–1/9/2009
Arnprior, monthly	6100345	T	°C	1/1/72–1/12/05	5.70	6.80	11.41	−1.27	−0.23	42.10	1/1/1959–9/9/1999
Russell, monthly	6107247	T	°C	1/1/72–1/12/05	6.32	7.20	10.98	−1.20	−0.25	40.30	1/1/1954–1/12/2006
Ottawa, CDA, monthly	6105976	P	mm	1/1/72–1/12/05	76.16	73.10	34.55	0.29	0.64	193.60	1/1/1889–1/12/06
Ottawa, Airport, monthly	6106000	P	mm	1/1/72–1/12/05	78.18	73.20	34.35	0.65	0.64	223.20	1/1/1938–1/12/06
Angers, QC, monthly	7030170	P	mm	1/1/72–1/12/05	82.11	76.00	39.17	0.58	0.74	209.30	1/1/1962–1/9/2009
Arnprior, monthly	6100345	P	mm	1/1/72–1/12/05	66.87	65.50	32.12	−0.19	0.50	174.00	1/1/1959–9/9/1999
Russell, monthly	6107247	P	mm	1/1/72–1/12/05	80.12	75.80	35.71	0.50	0.74	189.40	1/1/1954–1/12/2006
Ottawa, CDA, daily	6105976	P	mm	1/10/1972–9/9/1999	2.50	0.00	5.80	33.16	4.49	108.60	1/1/1889–1/12/06
Arnprior, daily	6100345	P	mm	1/10/1972–9/9/1999	2.20	0.00	5.36	26.16	4.21	76.00	1/1/1959–9/9/1999
Ottawa, CDA, daily	6105976	T <sub>max</sub>	°C	1/10/1972–9/9/1999	11.13	11.90	12.70	−0.90	−0.27	65.30	1/1/1889–1/12/06
Arnprior, daily	6100345	T <sub>max</sub>	°C	1/10/1972–9/9/1999	10.81	12.00	13.10	−0.90	−0.28	67.00	1/1/1959–9/9/1999
Ottawa, CDA, daily	6105976	T <sub>min</sub>	°C	1/10/1972–9/9/1999	1.65	2.60	11.82	−0.57	−0.47	59.80	1/1/1889–1/12/06
Arnprior, daily	6100345	T <sub>min</sub>	°C	1/10/1972–9/9/1999	0.80	2.00	12.21	−0.40	−0.54	63.00	1/1/1959–9/9/1999

T, air temperature; P, total precipitation; F, streamflow.

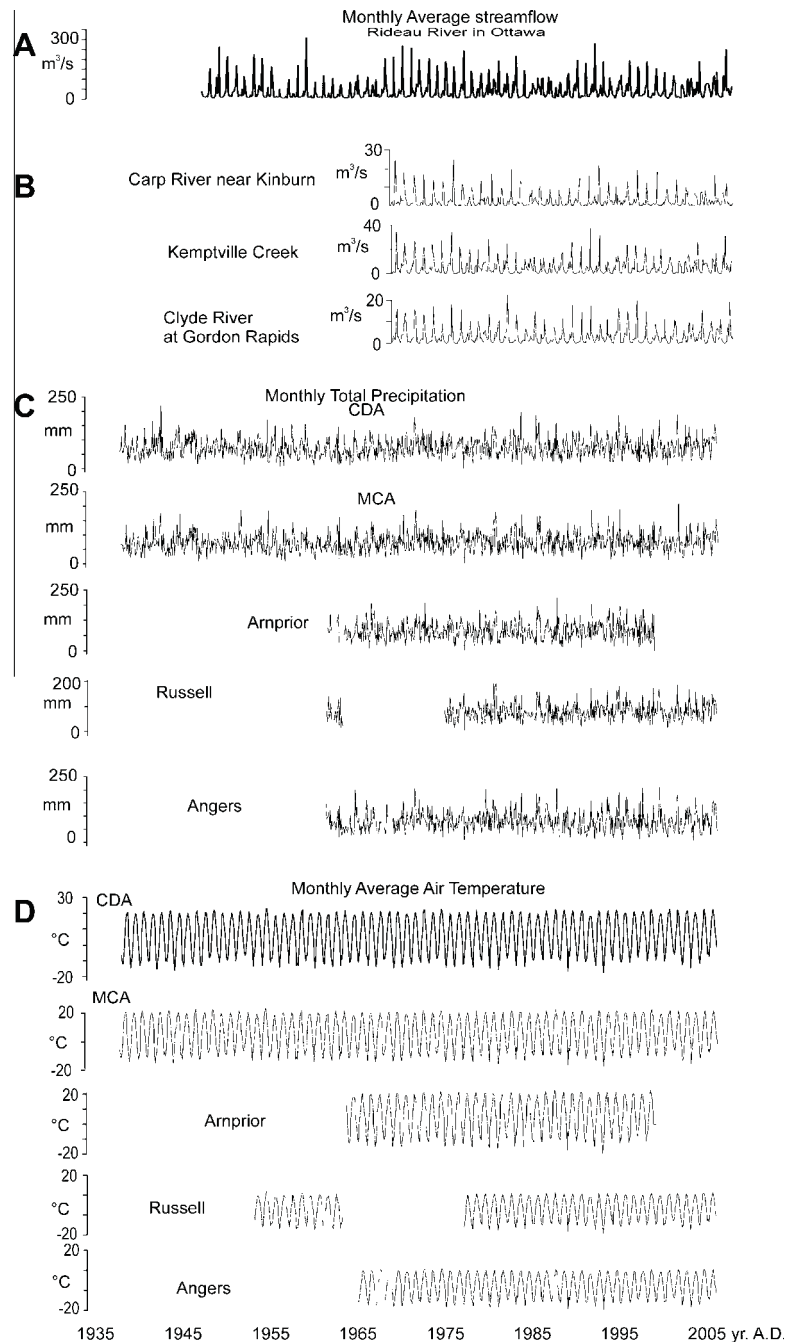


Fig. 4. Meteorological and hydrological monthly mean records from Eastern Canada.

2009). Thus, a log-normal transformation was required to provide a Gaussian distribution suitable for use in Morlet-wavelet-based CWT and XWT.

### 5.2. Long-term trends and pattern in climate and streamflow records

For the MK test and linear regression, climate records were downsampled into annual records for the years 1972–1999 for comparison with the streamflow records (Table 2). Neither of the annual mean records shows any significant trends with precipitation. Annual precipitation differences between both rural and both urban locations showed no significant trends (Table 2).

The relative difference between urban vs. rural warming was manifested by the difference between long-term trends in the ur-

ban Ottawa CDA, the semi-urban MCA records and the average of the three rural stations (Arnprior, Russell, and Angers) from 1979 to 1999. In support of the UHIE, the MK test and linear regression methods (Table 2) showed that the air temperature in the Ottawa city centre (Ottawa CDA station) warmed significantly ( $P \leq 0.01$ ) over this period, while at Ottawa's periphery (Ottawa MCA) this trend was less significant ( $P \leq 0.05$ ), and not significant ( $P > 0.05$ ) in the surrounding rural areas. The air temperature at the Russell station showed a weakly significant warming trend between 1979 and 2005 ( $P \leq 0.10$ ) when the population grew by 5–14% per year, compared to a normal growth of less than 3% per year. This indicates an urban heat island effect of  $0.022 \text{ } ^\circ\text{C}/\text{year}$  for Ottawa versus the surrounding rural areas (Table 2). The temperature trend difference between Ottawa CDA and the semi-urban MCA

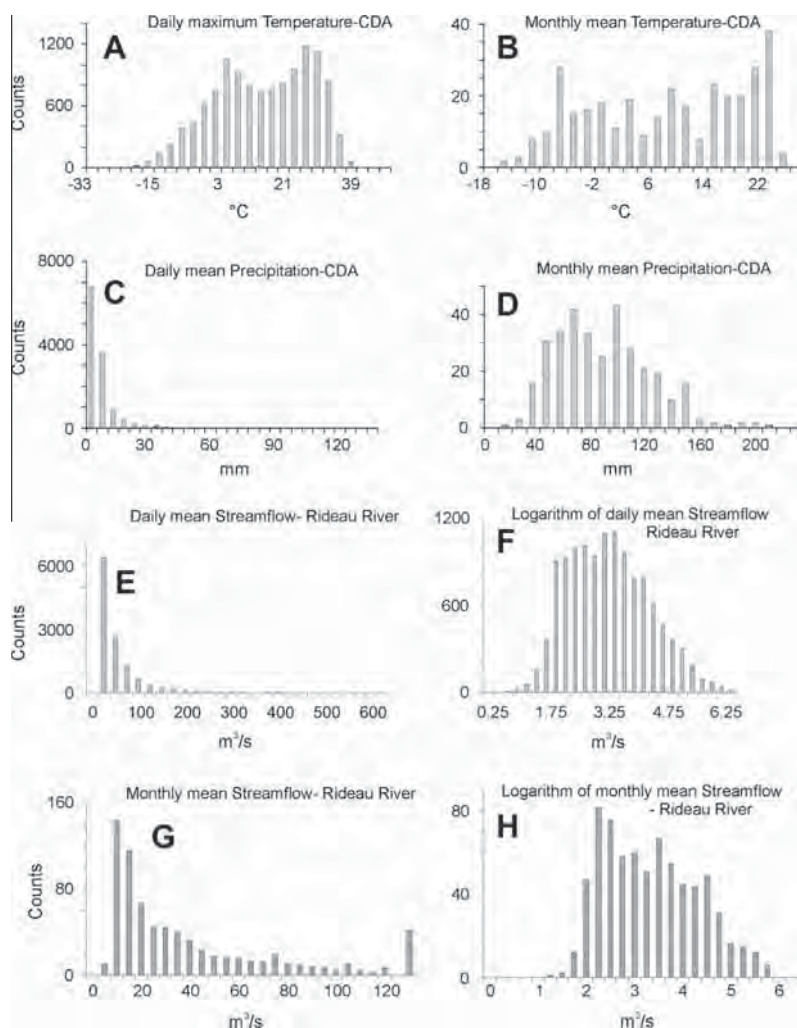


Fig. 5. Histograms of raw and logarithmic transformed meteorological and hydrological daily and monthly mean records.

Table 2

Linear regression and Mann–Kendall tests of long-term changes: trends and significance.

Time series	Data type	First year	Last year	<i>n</i>	Mann–Kendall trend		Regression $y = at + b$	
					Test Z	Signific.	Trend(a)	<i>r</i>
Rideau River	<i>F</i>	1972	2006	35	−0.37		−0.097	0.12
Carp River	<i>F</i>	1972	2006	35	−1.05		−0.01	0.22
Kemptville Creek	<i>F</i>	1972	2006	35	−0.97		−0.019	0.20
Clyde River	<i>F</i>	1972	2006	35	−0.48		−0.005	0.09
Ottawa CDA	<i>T</i>	1972	2005	34	2.58	0.010	0.042	0.50 <sup>*</sup>
Ottawa MCA	<i>T</i>	1972	2005	34	2.25	0.050	0.033	0.41 <sup>*</sup>
Arnprior	<i>T</i>	1972	1998	27	1.04		−0.027	0.18
Russell	<i>T</i>	1979	2005	27	1.71	0.100	0.034	0.37
Angers	<i>T</i>	1972	2005	34	0.27		0.004	0.05
Ottawa CDA	<i>P</i>	1972	2005	34	−0.47		−0.102	0.12
Arnprior	<i>P</i>	1972	1998	27	0.71		0.124	0.17

*T*, annual mean air temperature; *P*, total daily precipitation, annual average; *F*, annual average streamflow; *n*, record length (in years); *r*–Pearson correlation coefficient.  
<sup>\*</sup> Level of significance >0.05.

station is minor (Table 2), likely due to increased construction and expansion activities around the airport between 1980 and 2000. The trend of annual streamflow records in both rural and urban hydrological stations was found to be weakly decreasing, but not significant ( $P > 0.05$ ) using either MK tests or linear regression analysis (Table 2).

Cross-plots and the  $R^2$  values of linear regressions of climate vs. streamflow (Fig. 6) indicate a significant dependency of streamflow

records on precipitation, but not on temperature. This correlation was more significant in the urban data than in the rural data (Fig. 6C and D). Thus, relative wet periods over several years led to long-term increases in the streamflow over the same period, particularly for the Rideau River. The Rideau River transports more water and has a larger drainage area than the smaller Carp River and may thus be more responsive to larger amounts of precipitation.

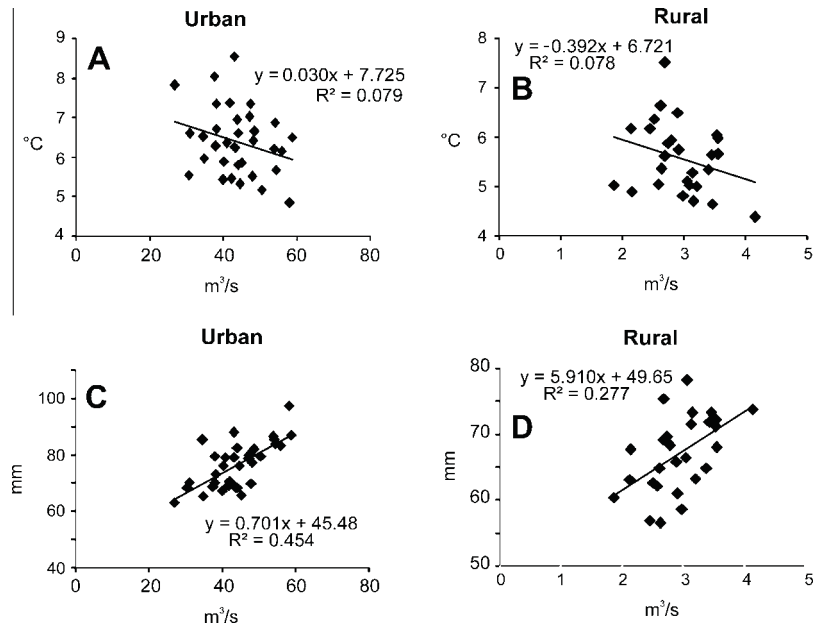


Fig. 6. Linear regression between mean annual meteorological and hydrological records from urban and rural areas with coefficient of determination ( $R^2$ ).

The rural area around the Carp River may be capable of releasing precipitation water through groundwater runoff more effectively. Alternatively, the weaker streamflow of the Carp River allows for a larger proportion of precipitation water to be removed from the streamflow by evaporation over the long term. However, this study indicates that the influence of precipitation on streamflow is the same in urban as in rural settings.

5.3. Impact of urban heat island effect on annual streamflow patterns

The CWT extracted the strongest signal for an annual amplitude in the 320–400 day waveband for all records from the urban Rideau River–Ottawa CDA and rural Carp River–Arnprior streamflow/weather station pairs (Fig. 7). Annual amplitude fluctuations were greater in the Rideau River with fluctuations of  $\approx 20 \text{ m}^3/\text{s}$  in years with a smaller difference between the annual snowmelt-induced discharge peak and base flow, and of  $\approx 50 \text{ m}^3/\text{s}$  in years when this difference was greater (Fig. 7A). The rural Carp River showed stronger relative fluctuations, as evidenced by the logarithmic-normalized record (Fig. 7B). Nevertheless, the temporal pattern was similar (Fig. 7A and B) with general negative trends. However, based on the MK test and Pearson correlation of both logarithmic normalized and non-normalized records, the annual amplitude trend was only significant in the rural Carp River (Table 3). This indicates that the contrast between high snowmelt-induced discharge and low baseflow significantly decreased over the study period, by either an increase in background flow or a decline in maximum flow.

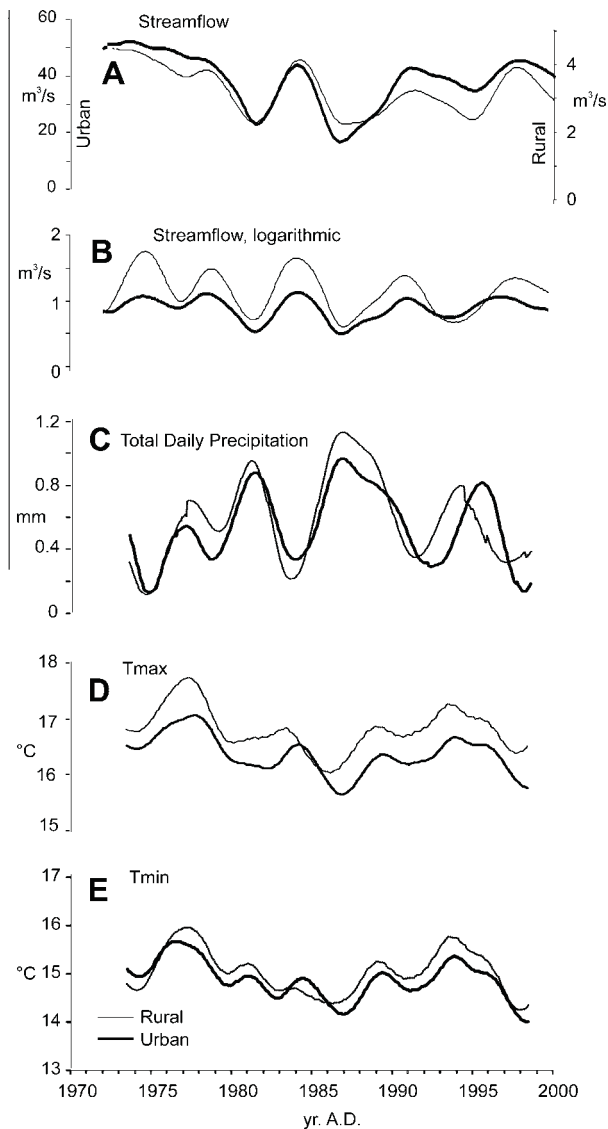
The annual amplitudes of the climate records showed similar patterns between rural and urban records, with obvious negative trends in  $T_{\min}$  and  $T_{\max}$  (Fig. 7C–E). The annual amplitude of variability in total precipitation from 1973 to 1999 showed similar fluctuations, but no significant trends in either rural or urban areas (Fig. 7C). Both locations had simultaneous years of relatively constant precipitation and years that were characterized by large precipitation amplitudes (i.e. swings between rainy and dry seasons).

The annual amplitudes of daily  $T_{\min}$  and  $T_{\max}$  showed a distinctly different pattern between rural and urban areas (Fig. 7D and E). However, the MK test and Pearson correlations indicated that only the negative trends in the urban setting are significant

(Table 3). Years with relatively high and low annual temperature swings were still well correlated, but Ottawa’s urban station (Ottawa CDA) showed only a slight linear decreasing trend from 1973 to 1999, indicating that the summer-winter temperature contrast became weaker over the years. This pattern is typical for the urban heat island effect (e.g., Karl et al., 1988).

The annual amplitude fluctuations of streamflow and precipitation were significantly correlated in both rural and urban areas (Fig. 8A and D). Thus, higher amounts of snowfall/rain lead to higher amplitudes of the snowmelt induced annual discharge cycle. This also indicates that local precipitation measurements describe the overall precipitation pattern and its influence on the drainage system of the streamflow of interest well in the study region. Urban areas with more variable precipitation over short distances may require more than “pinpoint” measurements to illustrate urban or rural precipitation patterns. In contrast, only the urban annual  $T_{\min}$  and  $T_{\max}$  amplitudes were significantly correlated ( $r^2 > 0.34$  and  $0.55$ , respectively) with annual variability in streamflow (Fig. 8E and F).

The phase difference of the annual waveband signal between urban and rural records shows a shift of  $\approx -0.017$  radians, indicating that the annual streamflow cycle of the Carp River lags  $\approx 1$  day behind the Rideau River in Ottawa (Fig. 9). There is an abrupt drop to  $\approx -0.8$  radians ( $\approx 1.5$  months) in 1988/1989, with a recovery to  $\approx -0.2$  in 1990 (Fig. 9A). The phase difference in precipitation between both settings is even higher:  $\pm 1$  radians on average, equivalent to  $\pm 2.5$  months, indicating a strong local variability of the occurrence time of precipitation events (Fig. 9B). Neither phase differences in streamflow nor precipitation showed any significant trend (Table 3). In contrast, the phase difference of the annual cycle between urban and rural areas developed from  $\approx 0$  to  $\approx -0.03$  ( $\approx 2$  days) for daily  $T_{\min}$  (Fig. 9C) and to  $\approx -0.013$  ( $\approx 1$  day) for daily  $T_{\max}$  (Fig. 9D), both of which were statistically significant (3). Moreover, there was a significant positive correlation between phase differences in streamflow and  $T_{\min}$ , indicating that an earlier annual warming in the late 1980s in Ottawa vs. the rural areas, may have resulted in an early onset of snow-melt induced high discharge in the Rideau River compared to the Carp River, where the threshold for ice-breakup was not reached for several weeks after the Rideau River.



**Fig. 7.** Amplitudes of annual (320–400 days) waveband of urban (Ottawa CDA, Rideau River) and rural (Arnprior, Carp River) meteorological and streamflow records.

**Table 3**

Mann–Kendall test of trends in annual amplitudes and phase of streamflow and climate records.

Time series	Data					MK trend		Regression $y = at + b$		Regression $\ln(y) = at + b$	
	Parameter	type	First year	Last year	$n$	Test Z	$\alpha$	Trend(a)	$r$	Trend(a)	$r$
RR	Amplitude	F	1974	1997	24	-1.36		-0.318	0.24	-0.002	0.09
CR	Amplitude	F	1974	1997	24	-1.71	0.1	-0.039	0.43**	-0.018	0.4**
CDA	Amplitude	P	1974	1997	24	0.72		0.006	0.17		
CDA- $T_{\min}$	Amplitude	T	1974	1997	24	-1.31		-0.018	0.39*		
CDA- $T_{\max}$	Amplitude	T	1974	1997	24	-1.41		-0.020	0.37*		
AR	Amplitude	P	1974	1997	24	-0.12		0.003	0.08		
AR- $T_{\min}$	Amplitude	T	1974	1997	24	-0.77		-0.010	0.16		
AR- $T_{\max}$	Amplitude	T	1974	1997	24	-0.62		-0.015	0.26		
RR-CR	Phase lag	F	1974	1997	24			-0.0390	0.17	0.0002	0.00
CDA-AR	Phase lag	P	1974	1997	24			-0.0185	0.31		
CDA-AR- $T_{\min}$	Phase lag	T	1974	1997	24			-0.0002	0.25		
CDA-AR- $T_{\max}$	Phase lag	T	1974	1997	24			-0.0002	0.45**		

$T$ , air temperature;  $P$ , total precipitation;  $F$ , streamflow;  $n$ , record length (in years); RR, Rideau River; CR, Carp River; CDA, Ottawa Canadian Department of Agriculture station; AR, Arnprior station;  $r$ -Pearson correlation coefficient.

\* Level of significance  $\alpha > 0.1$ .

\*\* Level of significance  $\alpha > 0.05$ .

## 6. Linkages in long-term and annual streamflows and climates

The long-term trends and correlations in the climate and hydrological records indicate a set of potential influences of UHIE on streamflow. However, the trends that were found are not related to the UHIE, but rather to regional or global variability. An influence of UHIE on long-term (multi-year) streamflow variability was not evident from our results. This is probably due to the suppression of the effects of local urban warming on precipitation by more influential, non-UHIE-mediated interannual changes, for example in precipitation, evaporation, and groundwater flow. For example, there was no significant relationship between long-term urban warming and mean streamflow discharge. The streamflow is, in the long-term, most strongly dependent on the precipitation pattern, which is not significantly impacted by the UHIE. In urban-heat-island-affected areas streamflow maxima peaked at lower values, but the large streamflows during the peak season were of longer duration than in rural areas. For example, large winter and summer flooding events occurred after high precipitation, whereas dry winters and wet summers were characterized by fewer and less extensive annual flooding events. The most significant detected impact of UHIE on streamflow occurred within the annual cycle, where the amplitude of temperature changes (both daily  $T_{\min}$  and  $T_{\max}$ ) were significantly correlated to streamflow amplitude changes, conferring the possibility of transfer functions.

The transfer function between temperature  $T$  ( $^{\circ}\text{C}$ ), and streamflow  $F$  ( $\text{m}^3/\text{s}$ ), was  $\Delta T/\text{year} \approx 0.025 \Delta F/\text{year}$ , or vice versa,  $\Delta F/\text{year} \approx 40 \Delta T/\text{year}$ . Thus, a reduction of  $\approx 1^{\circ}\text{C}$  in the annual winter cooling (or summer warming) may lead to a  $\approx 40 \text{ m}^3/\text{s}$  reduction in streamflow during the snowmelt induced annual discharge cycle. Over the last few decades the drop in the annual temperature amplitude at Ottawa CDA has been  $\approx 0.02^{\circ}\text{C}/\text{year}$  (Fig. 7D and E), which has resulted in a cumulative drop of  $\approx 0.8^{\circ}\text{C}$  over the last 40 years (Prokoph and Patterson, 2004). Based on the above relationship, this corresponds to a  $\approx 32 \text{ m}^3/\text{s}$  decrease in streamflow of the annual spring flood over the past four decades. This is a significant drop ( $>15\%$ ), considering that on average, the annual streamflow increases from an average baseflow of  $\approx 55 \text{ m}^3/\text{s}$  to over  $200 \text{ m}^3/\text{s}$  during the annual flood season; for the Rideau River, this season lasts one to two weeks. However, the absence of any significant correlation between air temperature and streamflow in the rural areas surrounding Ottawa underlines the impact and importance of the UHIE and the models used to detect it.



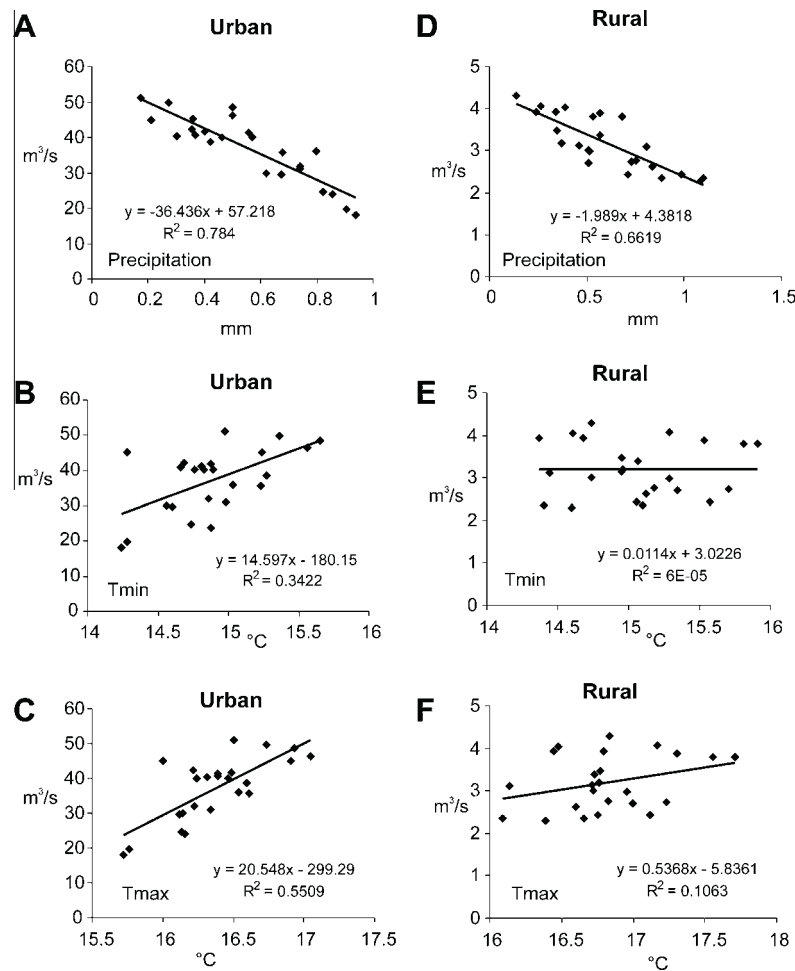


Fig. 8. Linear regression between annual amplitudes of meteorological and hydrological data from urban and rural areas with coefficient of determination ( $R^2$ ).

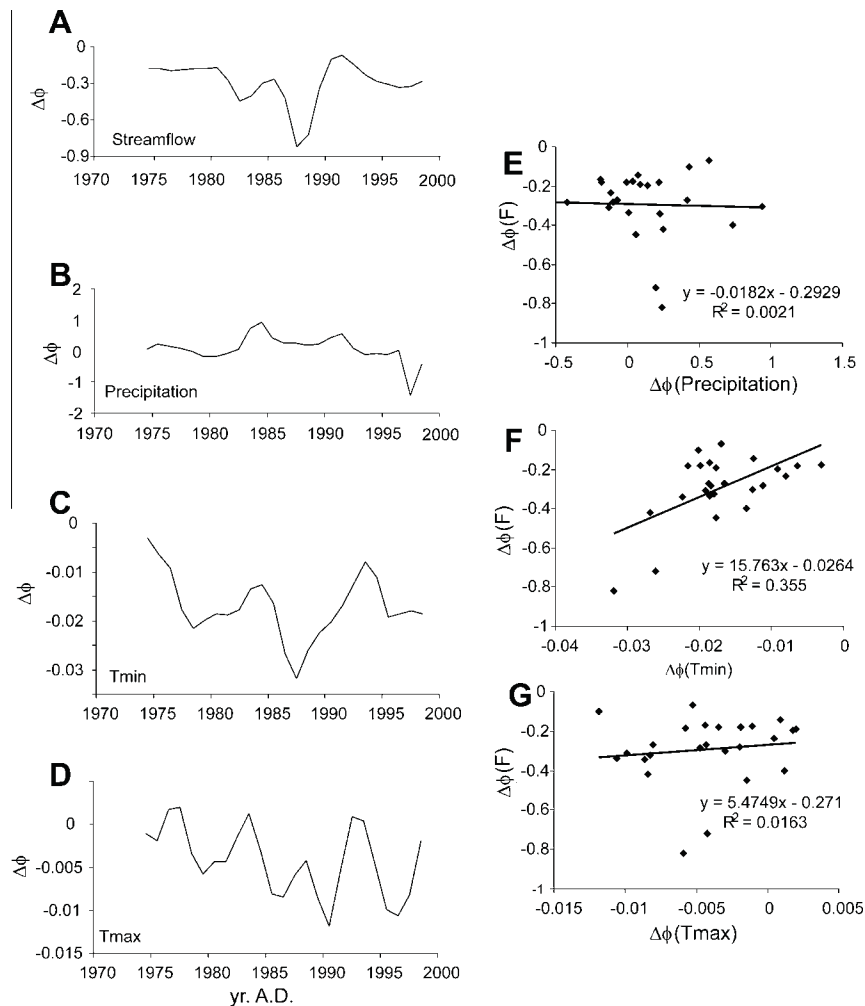
It was also found that a 1-day difference exists between urban and rural areas in the time lags between the onsets of the annual temperature cycle and streamflow cycle (Fig. 9). It is possible that this time-lag difference is caused by only a few extraordinary events during the late 1980s; more tests on longer datasets are required for stronger verification.

Previous studies have used combinations of wavelet analysis, regression analysis, or the Mann–Kendall test to explore waveband specific signals (e.g., Burn et al., 2008; Coulibaly, 2006). However, most of the previous studies focused on the relationship that exists between streamflow and long-term patterns, such as those seen with the North Atlantic Oscillation, or the sunspot and El Niño cycles (e.g., Anctil and Coulibaly, 2004; Adamowski et al., 2009). As well, no studies to date have used both wavelet analysis and the Mann–Kendall test together to explore the influence of climate on streamflow. This study determined, for the first time, in a quantitative manner, the impact of the urban heat island effect (i.e., urban warming) on the streamflow of rivers that flow through the center of their associated city. This study by no means exhausted all possible UHIE influences on river discharge. Including more climate related parameters and more study sites would provide a wider understanding. Moreover, higher-resolution records (e.g., hourly) may also detect diurnal dependency that could be useful for decision-makers involved with lock openings, waste overflow management and other issues. The methods used in this study (MK test, CWT and XWT) permit waveband-specific extraction and analysis of time-scale dependent changes that can provide

information beyond purely time-domain-averaged analyses such as those which use spectral techniques, or frequency-domain averaging analysis (e.g., smoothing and linear analysis). The use of time-independent correlation coefficients after the extraction stage already gives a good understanding of the UHIE, in particular for relatively short datasets such as the ones used in this study. For longer records the determination of time-dependent significance and coherency levels would be useful in tracing the development of an UHIE over time (e.g., a village growing into a large industrial city).

## 7. Conclusions

In this study, long-term (i.e. multi-annual) flow patterns were not significantly influenced by urban warming. However, annual flow patterns were significantly affected in the urban (vs. rural) area. The annual amplitude in the urban Rideau River streamflow in Ottawa correlated positively with the annual air temperature amplitude (i.e. less severe annual flooding with a decreasing winter/summer temperature contrast), whereas such a relationship did not occur at the rural stations. In the examples explored in this study, the annual flow pattern was strongly characterized by the magnitude and exact timing of the snowmelt-related high streamflow. Such snowmelt-induced annual discharge cycles are typical for most mid- to high-latitude and alpine streams in the northern hemisphere, which frequently pass through urban centers.



**Fig. 9.** Urban–Rural phase-lag differences in the annual waveband: (A–D) Records of phase-lag differences in the annual waveband of meteorological and hydrological data, note: 0.5 radians are equivalent to 1 month, and 0.017 radians are equivalent to 1 day. (E–F) Linear regression between phase-lag changes between urban and rural records of streamflow versus meteorological records, including coefficient of determination ( $R^2$ ).

This study also found that the precipitation pattern was not affected by urban heat island warming. The warming from the UHIE, especially during winter months, may reduce the severity of the annual spring flood event in mid-to-high latitudinal continental settings. Further studies could focus on the development of transfer functions between monthly or higher resolution climate records and corresponding streamflow measurements, as well as include other records that may be of importance to city planners and policy-makers in quantifying and projecting streamflow patterns (e.g., daily or monthly minimum or monthly flow) in a warming environment. For example, paired meteorological–streamflow measurements could be assigned for upstream and downstream regions of the urban center to evaluate the persistence and resilience of streamflow to the UHIE. The methodology presented permits a wide range of bandwidth-dependent climate record extractions and their subsequent analysis (e.g. as with XWT) and comparison between urban and rural data sources. A method could be developed to determine the degrees of freedom of bandwidth, mother wavelet and time resolution dependent wavelet amplitudes to better test the significance of potential relationships.

#### Acknowledgements

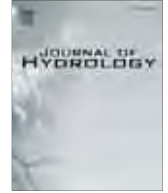
This study was funded by a National Science and Engineering Research Council of Canada Discovery Grant held by Jan

Adamowski. The authors would like to thank P. Larson from the Rideau River Valley Conservation authority for providing streamflow data, and two anonymous reviewers for their suggestions.

#### References

- Abdul Aziz, O.I., Burn, D.H., 2006. Trends and variability in the hydrological regime of the Mackenzie River Basin. *J. Hydrol.* 319 (1–4), 282–294. <http://dx.doi.org/10.1016/j.jhydrol.2005.06.039>.
- Adamowski, J.F., 2008. Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. *J. Hydrol.* 353 (3–4), 247–266. <http://dx.doi.org/10.1016/j.jhydrol.2008.02.013>.
- Adamowski, K., Prokoph, A., Adamowski, J., 2009. Development of a new method of wavelet aided trend detection and estimation. *Hydrol. Process.* 23 (18), 2686–2696. <http://dx.doi.org/10.1002/hyp.7260>.
- Al-Smadi, A., 2005. Tests for Gaussianity of a stationary time series. *World Acad. Sci. Eng. Technol.* 10, 120–124.
- Anctil, F., Coulibaly, P., 2004. Wavelet analysis of the interannual variability in Southern Quebec streamflow. *J. Clim.* 17 (1), 163–173. [http://dx.doi.org/10.1175/1520-0442\(2004\)017<0163:WAOTIV>2.0.CO](http://dx.doi.org/10.1175/1520-0442(2004)017<0163:WAOTIV>2.0.CO).
- Arnfield, A.J., 2003. Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island. *Int. J. Climatol.* 23, 1–26. <http://dx.doi.org/10.1002/joc.859>.
- Burn, D.H., Fan, L., Bell, G., 2008. Identification and quantification of streamflow trends on the Canadian Prairies. *Hydrol. Sci.* 53 (3), 538–549. <http://dx.doi.org/10.1623/hysj.53.3.538>.
- Coulibaly, P., 2006. Spatial and temporal variability of Canadian seasonal precipitation (1900–2000). *Adv. Water Resour.* 29 (12), 1846–1865. <http://dx.doi.org/10.1016/j.advwatres.2005.12.013>.
- Davis, J.C., 1986. *Statistics and Data Analysis in Geology*. Wiley, New York, 646pp.

- Douglas, E.M., Vogel, R.M., Kroll, C.N., 2000. Trends in floods and low flows in the United States: impact of spatial correlation. *J. Hydrol.* 240 (1), 90–105. [http://dx.doi.org/10.1016/S0022-1694\(00\)00336-X](http://dx.doi.org/10.1016/S0022-1694(00)00336-X).
- Gilbert, R.O., 1987. *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold, New York, 320pp.
- Granato, G.E., 2009. *Computer Programs for Obtaining and Analyzing Daily Mean Streamflow Data from the U.S. Geological Survey National Water Information System Web Site: U.S. Geological Survey Open-File Report 2008-1362*, 123pp.
- Hinkel, K.M., Nelson, F.E., Klene, A.E., Bell, J.H., 2003. The urban heat island in winter at Barrow, Alaska. *Int. J. Climatol.* 23, 1889–1905. <http://dx.doi.org/10.1002/joc.971>.
- Howard, L., 1833. *Climate of London Deduced from Meteorological Observations*. third ed., vol. 1. Harvey and Darton, London, 348pp.
- Jury, M.R., Enfield, D.B., Méléce, J., 2002. Tropical monsoons around Africa: stability of El Niño–Southern Oscillation associations and links with continental climate. *J. Geophys. Res.* 107, 3151–3167.
- Karaca, M., Tayanç, M., Toros, H., 1995. Effects of urbanization on climate of Istanbul and Ankara. *Atmos. Environ.* 29 (23), 3411–3421. [http://dx.doi.org/10.1016/1352-2310\(95\)00085-D](http://dx.doi.org/10.1016/1352-2310(95)00085-D).
- Karl, T.R., Diaz, H.F., Kukla, G., 1988. Urbanization: its detection and effect in the United States climate record. *J. Clim.* 1 (11), 1099–1123. [http://dx.doi.org/10.1175/1520-0442\(1988\)001<1099:UIDAEi>2.0.CO;2](http://dx.doi.org/10.1175/1520-0442(1988)001<1099:UIDAEi>2.0.CO;2).
- Kendall, M.G., 1975. *Rank Correlation Methods*, forth ed. Charles Griffin, London.
- Kinouchi, T., Yagi, H., Miyamoto, M., 2007. Increase in stream temperature related to anthropogenic heat input from urban wastewater. *J. Hydrol.* 335 (1–2), 78–88. <http://dx.doi.org/10.1016/j.jhydrol.2006.11.002>.
- Krakauer, N.Y., Fung, I., 2008. Is streamflow increasing? Trends in the coterminous United States. *Hydrol. Earth Syst. Sci. Discuss.* 5, 785–810.
- Labat, D., 2008. Wavelet analysis of the annual discharge records of the world's largest rivers. *Adv. Water Resour.* 31 (1), 109–117.
- Lettenmaier, D.P., Wood, E.F., Wallis, J.R., 1994. Hydro-climatological trends in the continental United States, 1948–88. *J. Clim.* 7 (4), 586–607. [http://dx.doi.org/10.1175/1520-0442\(1994\)007<0586:HCTITC>2.0.CO;2](http://dx.doi.org/10.1175/1520-0442(1994)007<0586:HCTITC>2.0.CO;2).
- Li, Q., Zhang, H., Liu, X., Huang, J., 2004. Urban heat island effect on annual mean temperature during the last 50 years in china. *Theor. Appl. Climatol.* 79 (3–4), 165–174. <http://dx.doi.org/10.1007/s00704-004-0065-4>.
- Lin, C.Y., Chen, W.C., Chang, P.L., Sheng, Y.F., 2011. Impact of the urban heat island effect on precipitation over a complex geographic environment in Northern Taiwan. *J. Appl. Meteorol. Climatol.* 50 (2), 339–353. <http://dx.doi.org/10.1175/2010JAMC2504.1>.
- Manley, G., 1958. On the frequency of snowfall in metropolitan England. *Q. J. R. Meteorol. Soc.* 84, 70–72. <http://dx.doi.org/10.1002/qj.49708435910>.
- Mann, S., Haykin, S., 1992. Adaptive “chirplet” transform: an adaptive generalization of the wavelet transform. *Opt. Eng.* 31 (6), 1243–1256.
- Maraun, D., Kurths, J., 2004. Cross wavelet analysis: significance testing and pitfalls. *Nonlin. Process. Geophys.* 11, 505–514.
- Montávez, J.P., Rodríguez, A., Jiménez, J.L., 2000. A study of the heat island of Granada. *Int. J. Climatol.* 20, 899–911. [http://dx.doi.org/10.1002/1097-0088\(20000630\)20:8<899::AID-JOC433>3.0.CO;2-I](http://dx.doi.org/10.1002/1097-0088(20000630)20:8<899::AID-JOC433>3.0.CO;2-I).
- Morlet, J., Arens, G., Fargeau, I., Giard, D., 1982a. Wave propagation and sampling theory – Part I: complex signal and scattering in multilayered media. *Geophysics* 47 (2), 203–221. <http://dx.doi.org/10.1190/1.1441328>.
- Morlet, J., Arens, G., Fargeau, I., Giard, D., 1982b. Wave propagation and sampling theory – Part II: sampling theory and complex waves. *Geophysics* 47 (2), 222–236. <http://dx.doi.org/10.1190/1.1441329>.
- Moshin, T., Gough, W.A., 2012. Characterization and estimation of urban heat island at Toronto: impact of the choice of rural sites. *Theoret. Appl. Climatol.* 108 (1–2), 105–117. <http://dx.doi.org/10.1007/s00704-011-0516-7>.
- Nakken, M., 1999. Wavelet analysis of rainfall-runoff variability isolating climatic from anthropogenic patterns. *Environ. Model. Softw.* 14 (4), 283–295. [http://dx.doi.org/10.1016/S1364-8152\(98\)00080-2](http://dx.doi.org/10.1016/S1364-8152(98)00080-2).
- Oke, T.R., 1973. City size and the urban heat island. *Atmos. Environ.* 7 (8), 769–779. [http://dx.doi.org/10.1016/0004-6981\(73\)90140-6](http://dx.doi.org/10.1016/0004-6981(73)90140-6).
- Oke, T.R., Maxwell, G.B., 1975. Urban heat island dynamics in Montreal and Vancouver. *Atmos. Environ.* 9 (2), 191–200. [http://dx.doi.org/10.1016/0004-6981\(75\)90067-0](http://dx.doi.org/10.1016/0004-6981(75)90067-0).
- Prokoph, A., Barthelme, F., 1996. Detection of nonstationarities in geological time series: wavelet transform of chaotic and cyclic sequences. *Comput. Geosci.* 22 (10), 1097–1108. [http://dx.doi.org/10.1016/S0098-3004\(96\)00054-4](http://dx.doi.org/10.1016/S0098-3004(96)00054-4).
- Prokoph, A., Patterson, R.T., 2004. Application of wavelet and regression analysis in assessing temporal and geographic climate variability: Eastern Ontario, Canada as a case study. *Atmosphere–Ocean* 42 (3), 201–212. <http://dx.doi.org/10.3137/ao.420304>.
- Rioul, O., Vetterli, M., 1991. Wavelets and signal processing. *IEEE Signal Process. Mag.* 8 (4), 14–38. <http://dx.doi.org/10.1109/79.91217>.
- Salmi, T., Määttä, A., Anttila, P., Ruoho-Airola, T., Amnell, T., 2002. Detecting Trends of Annual Values of Atmospheric Pollutants by the Mann-Kendall Test and Sen's Slope Estimates – The Excel Template Application MAKESENS. *Publications on Air Quality*. No. 31. Finnish Meteorological Institute, Finland, 35pp. <[http://www.ilmanlaatu.fi/ilmansaasteet/julkaisu/pdf/MAKESENS-Manual\\_2002.pdf](http://www.ilmanlaatu.fi/ilmansaasteet/julkaisu/pdf/MAKESENS-Manual_2002.pdf)> (accessed 14.02.13).
- Schaeffli, B., Zehe, E., 2009. Hydrological model performance and parameter estimation in the wavelet-domain. *Hydrol. Earth Syst. Sci.* 13 (10), 1921–1936.
- Shepherd, J.M., Burian, S.J., 2003. Detection of urban-induced rainfall anomalies in a major coastal city. *Earth Interact.* 7, 1–17. [http://dx.doi.org/10.1175/10873562\(2003\)007<0001:DOUIRA>2.0.CO;2](http://dx.doi.org/10.1175/10873562(2003)007<0001:DOUIRA>2.0.CO;2).
- Stewart, I.D., 2000. Influence of meteorological conditions on the intensity and form of the urban heat island effect in Regina. *Can. Geogr.* 44 (3), 271–285. <http://dx.doi.org/10.1111/j.1541-0064.2000.tb00709.x>.
- Torrence, C., Compo, G.P., 1998. A practical guide to wavelet analysis. *Bull. Amer. Meteorol. Soc.* 79 (1), 61–78. [http://dx.doi.org/10.1175/1520-0477\(1998\)079<0061:APGTWA>2.0.CO;2](http://dx.doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2).
- Ware, D.M., Thomson, R.E., 2000. Interannual to multidecadal timescale climate variations in the Northeast Pacific. *J. Clim.* 13 (18), 3209–3220. [http://dx.doi.org/10.1175/1520-0442\(2000\)013%3C3209:ITMTCV%3E2.0.CO;2](http://dx.doi.org/10.1175/1520-0442(2000)013%3C3209:ITMTCV%3E2.0.CO;2).
- Yague, C., Zurita, E., 1991. Statistical Analysis of the Madrid Urban Heat Island. *Atmos. Environ.* 25B (3), 327–332. [http://dx.doi.org/10.1016/0957-1272\(91\)90004-X](http://dx.doi.org/10.1016/0957-1272(91)90004-X).
- Yalcin, T., Yetemen, O., 2009. Local warming of groundwaters caused by the urban heat island effect in Istanbul, Turkey. *Hydrogeol. J.* 17 (5), 1247–1255. <http://dx.doi.org/10.1007/s10040-009-0474-7>.
- Yulianti, J.S., Burn, D.H., 1998. Investigating links between climatic warming and low streamflow in the prairies region of Canada. *Can. Water Resour. J.* 23 (1), 45–60. <http://dx.doi.org/10.4296/cwrj2301045>.
- Zhang, X., Harvey, K.D., Hogg, W.D., Yuzyk, R., 2001. Trends in Canadian streamflow. *Water Resour. Res.* 37 (4), 987–999. <http://dx.doi.org/10.1029/2000WR900357>.



# Quantifying the spatial temporal variability of annual streamflow and meteorological changes in eastern Ontario and southwestern Quebec using wavelet analysis and GIS



Jan Adamowski<sup>a,\*</sup>, Kaz Adamowski<sup>b</sup>, Andreas Prokoph<sup>c,d</sup>

<sup>a</sup> Department of Bioresource Engineering, McGill University, 21111 Lakeshore Road, Ste. Anne de Bellevue, QC H9X 3V9, Canada

<sup>b</sup> Department of Civil Engineering, University of Ottawa, Ottawa, ON K1N 6N5, Canada

<sup>c</sup> Department of Earth Sciences, Carleton University, Ottawa, ON K1S 5B6, Canada

<sup>d</sup> Speedstat, 19 Langstrom Crescent, Ottawa, ON K1G 5J5, Canada

## ARTICLE INFO

### Article history:

Received 23 June 2011

Received in revised form 21 April 2013

Accepted 14 June 2013

Available online 4 July 2013

This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Luis E. Samaniego, Associate Editor

### Keywords:

Streamflow

GIS

Climate change

Ontario

Quebec

Precipitation

## ABSTRACT

To make good policy decisions, knowledge of the trends and patterns in regional and local fluctuations in water resources and the influence of regional meteorological conditions are required. This study quantified the spatial–temporal variability of annual streamflow variations and meteorological changes in parts of eastern Canada. Time series and Geographic Information System (GIS) methods were used to evaluate the potential influence of low frequency climate change on annual spring flood onset and streamflow amplitude over a 24 year period. Continuous wavelet (CWT) and cross-wavelet transforms (XWT) were used to detect and extract temporal changes in the annual streamflow cycle amplitude, and to determine its relative phase shift to a standing 1-year sine wave through time. This was done to determine spatial–temporal differences and trends in the annual flood onset, and to determine the differences in the range between flood and background levels in 23 hydrological stations in eastern Canada. The study was carried out in the eastern Ontario/southwestern Quebec region of Canada because of the relatively high spatial density of hydrological stations from generally pristine waterways, as well as the high density of meteorological stations with uninterrupted recordings for a 24 year period from 1/1969 to 12/1992. This research indicates that over this 24 year period, the southwestern part of the region, along the US border, experienced an increase in temperature of up to 0.05 °C/year, and a decrease in precipitation of ~0.5 mm/month; the predominantly rural northwest region warmed only <0.02 °C/year and became <0.2 mm/year wetter. During the same time interval, the annual streamflow shifted on average to a ~20 days earlier flood onset, following the similar northwest–southeast trend of the meteorological records. In the northwest, along the Ottawa River, flooding occurred approximately 50 days earlier in the early 1990s than in the late 1960s, while it was less than 10 days earlier in the southeastern streams. The streamflows of the Ottawa and Rideau River, the main waterways that were studied, have generally lower annual streamflow amplitudes than the smaller and more pristine waterways that experience more seasonality, while the annual streamflow amplitude does not show any significant trend in time and space.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

Knowledge of the relationship between climate and water resource fluctuations is important for water resource predictions and management in a changing climate. The worldwide observational temperature records have revealed a global temperature increase of 0.3–0.7 °C in the last century (IPCC, 2007). In the northern hemisphere, most of this increase has been attributed to an increase in diurnal and annual temperatures, suggesting that conditions are actually ‘less cold’ rather than ‘warmer’ (e.g., Bonsal

et al., 2001). It has been estimated that southern Canada warmed between 0.5 and 1.5 °C through the 20th century due to several sources of anthropogenic warming, in particular greenhouse gas output (Bonsal et al., 2001; Zhang et al., 2000). In the same time interval, precipitation and streamflow patterns underwent significant, albeit regional, variations (Ehsanzadeh and Adamowski, 2010; Coulibaly, 2006; Coulibaly and Burn, 2005). For example, over the last 50 years, a large decrease in annual mean streamflow has been observed in British Columbia (Western Canada), which can be linked to a large increase in spring temperatures (Zhang et al., 2001).

The climate in eastern Ontario/southwestern Quebec is continental, with long, cold winters, short hot-humid summers, and slightly moderated by the effects of the Great lakes and the St. Lawrence River. The general climate pattern in the region is controlled

\* Corresponding author. Tel.: +1 5143987786.

E-mail addresses: [jan.adamowski@mcgill.ca](mailto:jan.adamowski@mcgill.ca) (J. Adamowski), [adamowski@eng.uottawa.ca](mailto:adamowski@eng.uottawa.ca) (K. Adamowski), [aprokocon@aol.com](mailto:aprokocon@aol.com) (A. Prokoph).



by the jet stream with predominant westerly winds, and to a weaker degree by El Niño's and the North Atlantic Oscillation (NAO), and in the summer by moisture from the Gulf of Mexico (e.g., Zhang et al., 2001). The climate in the metropolitan area of Ottawa, the area of focus in this study, is overprinted by urban heat island warming of  $\sim 1$  °C through the last century (Prokoph and Patterson, 2004). The streamflow pattern in and around Ottawa is characterized by strong spring flooding, with relatively constant low streamflows during the rest of the year (Adamowski, 2008). Occasionally, large rainfall events or warm spells in winter lead to additional flooding, providing generally weaker peaks in the streamflow than the main spring flood.

Several studies have incorporated both spatial and temporal information, and combined streamflow patterns (means and extremes) with climate records (Adamowski and Bocci, 2001; Coulibaly and Burn, 2005). The results of several spatial–temporal studies using Geographic Information Systems (GIS) in watersheds in Canada have found that trends differ significantly from region to region, or there may not be any detectable trends. While GIS based climate and water resource studies on a large scale are common for Canada, only a few have examined water resources in the Ottawa region. Those studies focused on water quality such as chemical composition (e.g., Telmer and Veizer, 1996), while other GIS studies focused on geological mapping in eastern Ontario, Canada (e.g., Couture et al., 2006).

The temporal trends in streamflow and climate are of a similar complexity. Ko and Cheng (2004) related meteorological data to hydrological data showing different response times of streamflow in a relatively small area in southern Ontario. In addition, there are only a few regional studies on potential climate–streamflow connections, such as on trends in the onset of floods in the western USA (e.g., Steward et al., 2005), or on changes in the amplitude of the annual streamflow cycles relative to precipitation in southern Canada (Adamowski et al., 2011). There are indications of a significant increase in precipitation and streamflow in the northeastern USA through the last century (e.g., Krakauer and Fung, 2008).

The combined spatial–temporal analysis of coeval meteorological and streamflow fluctuation are complicated by several issues. These include:

- (a) Dams and other man-made land-use changes influence streamflows independent of climate changes.

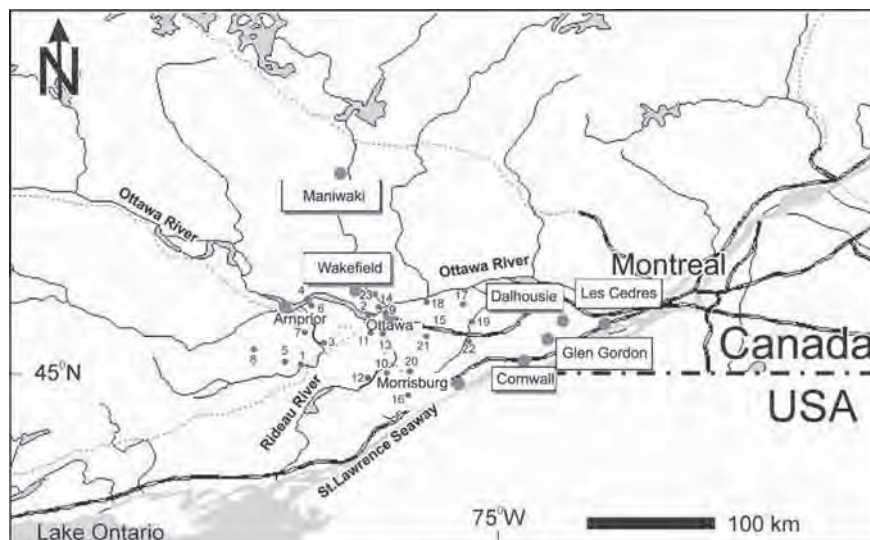
- (b) Coeval records from both meteorological and streamflow measurements do not exist in close proximity and over long time intervals.
- (c) Local meteorological records represent local conditions while streamflow patterns reflect conditions along the entire drainage system.
- (d) Local warming due to urbanization ('urban heat island effect', e.g., Karl et al., 1988) overprints regional and global climate patterns.
- (e) Only a few streamflow records are complete.

Thus, this may make it very difficult to draw solely data-based conclusions on future streamflow levels in Canada and their link to global climate change for both annual mean flow (e.g., Burn and Hag Elnur, 2002; Pilon and Yue, 2002; Adamowski and Bocci, 2001; Zhang et al., 2001), as well as streamflow extremes (e.g., Burn et al., 2010; Cunderlik and Burn, 2002).

In this study, changes in the annual streamflow amplitude and streamflow distributions were compared with changes in mean temperature and precipitation in eastern Ontario and southwestern Quebec, Canada, through the use of time series analysis techniques (trend, wavelet, cross-wavelet analysis) and spatial analysis using ArcGIS®. Eastern Ontario and southwestern Quebec have the advantage of having many, still pristine streamflows, with a relatively high density of meteorological and streamflow gauge stations where data has been recorded without interruptions for over 24 years (Fig. 1). Data was sampled from different locations in the same region, as well data representing surface hydrology and atmospheric conditions, and were combined and analyzed to allow for the interpretation of their spatial–temporal relationship. Knowledge of such relationships can help policy makers to mitigate water supply and flood problems based on knowledge of the regional meteorological trends and other temporal patterns. At the same time, it is possible to determine which hydrological regions are more susceptible to short and long term meteorological fluctuations.

## 2. Data

Complete monthly precipitation and temperature records from 9 stations in eastern Ontario and southwestern Quebec, for the time interval from January 1969 to December 1992, were



**Fig. 1.** Map eastern Ontario/southwestern Quebec, Canada: Red dots mark 23 streamflow gauge stations (for names and ID's see Table 1) and gray dots (except for Ottawa) mark meteorological stations used in this study.



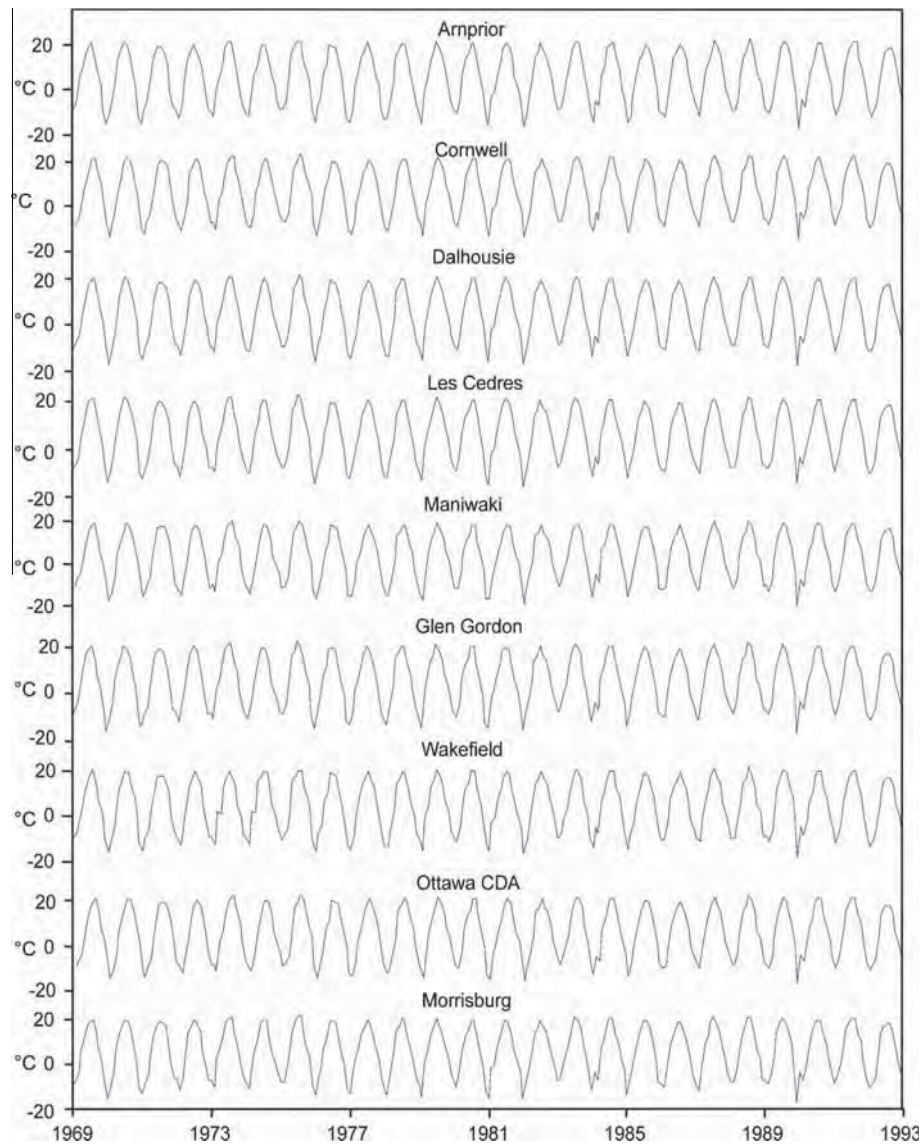


Fig. 2. Monthly air temperature measurements (in °C) from 9 stations in eastern Ontario/southwestern Quebec. For locations see Fig. 1.

downloaded from the Government of Canada Department of Environment website (<http://climate.weatheroffice.gc.ca>). About 20 additional meteorological stations are located in the region, but they have record gaps. The precipitation record for station Whitefield is an exception, with a 9 month gap during 1973. This record was kept for analysis because of its regionally relevant location north of Ottawa. This was deemed acceptable since the analysis covers a longer time interval (24 years), and precipitation does not have a strong seasonal signal like streamflow and temperature. In 1993, large budget cuts required the closure of most of the stations, thus restricting spatial analysis in this study to the time interval from 1969 to 1992. The monthly temperature record is dominated by the annual cycle with mean winter temperatures of  $\sim -12$  °C and summer temperatures of  $\sim 22$  °C (Fig. 2). Total monthly precipitation ranges from  $\sim 20$  to 200 mm in the region (Fig. 3), showing much less annual cyclicity and less regional correlation than the temperature records.

Monthly mean streamflow (discharge) records from 23 stations from the area at longitude  $W76.5^{\circ}$ – $W75.5^{\circ}$ , and latitude  $N44.5^{\circ}$ – $N45.5^{\circ}$ , were also made available by Environment Canada (<http://climate.weatheroffice.gc.ca>). These streamflow records show

average monthly streamflows ranging from  $\sim 1$  to  $1000$  m<sup>3</sup>/s, reflecting the different sizes of the drainage areas, topography, as well as watershed widths. Four stations are located along the Ottawa River with an increasing drainage area from west to east. The standard deviations of the streamflow are in accordance with the average streamflows, except for station 02LBO10N (Ottawa River at Cumberland), which is regulated by a series of dams, and used as an example of non-pristine conditions in this study (Fig. 4A). In all other stations, the annual variability is relatively constant compared to the streamflow average (Figs. 4B, 5). For example, station 02KFO09N (Ottawa River at Chats Falls), which is also regulated by a dam for a power station as well as the downriver located station Ottawa River at Britannia Bay show monthly-averaged streamflow patterns characteristic for pristine streamflows (Fig. 5). Nine of the stations have complete records (Table 1), whereas the other stations have missing records in the beginning, middle, or at the end of the 1969–1992 time interval. Missing records are due to: (i) instrumental and recording malfunctions that result in short-time gaps, mostly shorter than a week; or (ii) long-term gaps in records due to stopped measurements/recording caused by government spending cuts. The latter resulted in the

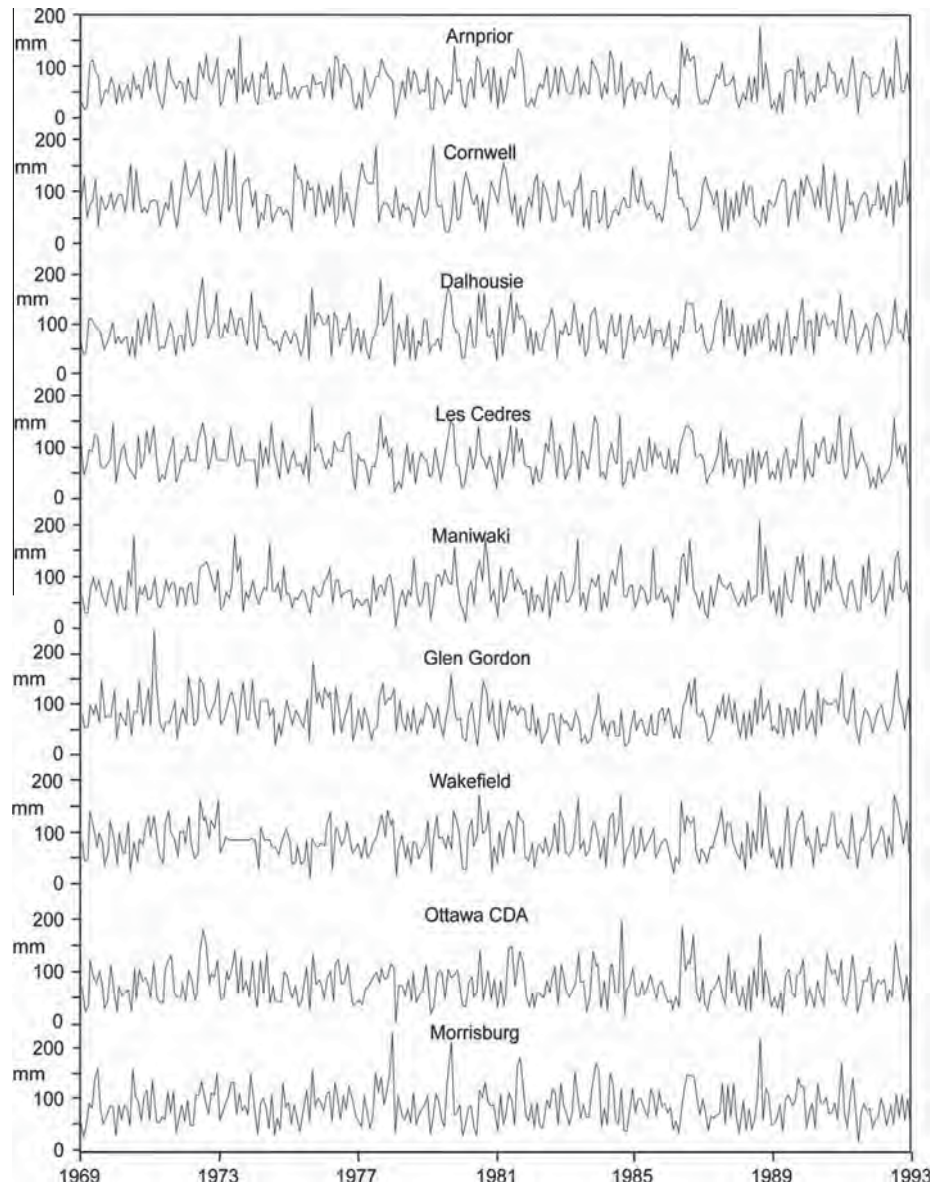


Fig. 3. Monthly total precipitation measurements (in mm) from 9 stations in eastern Ontario/southwestern Quebec. For locations see Fig. 1.

variable start and end of the records used in this analysis. In this study, all 23 stations have at least 8 years of uninterrupted streamflow records. In addition, all streamflow stations have complete records for 1986, which is used as a 'standard year' and used for normalization of the magnitudes and standard deviation streamflow between the stations.

### 3. Methods

#### 3.1. Analysis of streamflow records

Each streamflow record  $F$  was normalized using the annual standard deviation ( $S_{1986}$ ) and annual mean  $F_{1986}$  calculated from the 12 monthly records from 1986 to (a) relate trends and other temporal patterns consistently to the same year for all records, and (b) to adjust for the large differences in total discharge (streamflow) and variability (standard deviation) at the stations. The commonly used normalization utilizing mean and standard deviation of the entire dataset (e.g., Davis, 1986) was not used, because of potential bias due to uneven record lengths and missing

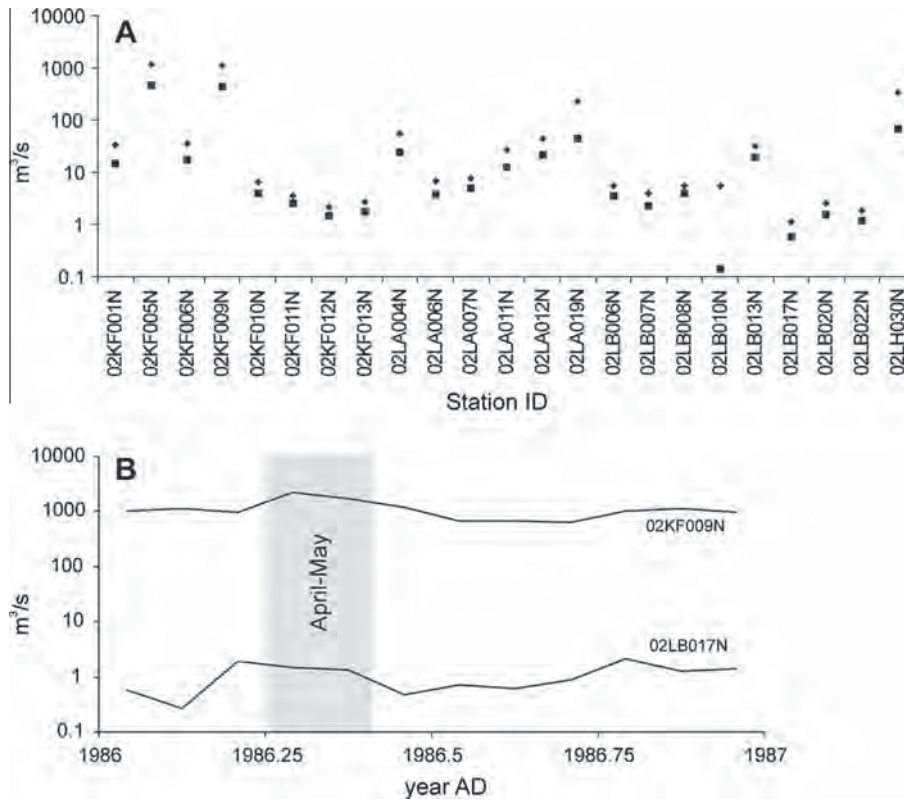
records. The normalized streamflow  $F_N$  for each station  $x$  at time  $t$  (Fig. 5) was calculated by:

$$F_N(x, t) = (F(x, t) - F_{1986}(x)) / S_{1986}(x) \quad (1)$$

Then, wavelet and cross-wavelet analysis were carried out on the normalized streamflow records to detect and extract the Fourier amplitudes and the phase in the annual (0.8–1.2 year) waveband, as well as the phase shift to a standing cosine wave with amplitude 1 and a wavelength of 12 months (= 1 year).

#### 3.2. Wavelet and cross-wavelet analysis

Wavelet analysis emerged as a filtering and data compression method in the 1980s (e.g., Morlet et al., 1982). Wavelet analysis transforms information from a 'depth' or 'time' domain into a spectral domain by using various shapes and sizes of short filtering functions, so called 'wavelets'. An advantage of wavelet over Fourier (spectral) analysis is that the wavelet coefficients are related to a specific period of time and frequency simultaneously (Fig. 6B), making it possible to trace the amplitude and phase of



**Fig. 4.** Streamflow statistics for 23 stations (for ID's see Table 1): (A) logarithmically scaled means of monthly streamflow (black diamonds) and their standard deviations (black squares) for 1969–1992; (B) range of monthly average streamflow records for 1986 based on the smallest (bottom) and largest streamflow (top) studied.

fluctuations of a specific wavelength through time. Single window spectral analysis provides a very accurate estimate of variance (or power) or Fourier amplitude, but is averaged over the discrete length of the analyzed dataset (Fig. 6A). As a result, changes of a wavelength specific Fourier amplitude through time cannot be measured.

The wavelet coefficients  $W$  of a time series  $x(s)$  are calculated by a simple convolution (Chao and Naito, 1995):

$$W_{\psi}(a, b) = \left(\frac{1}{a}\right) \int x(s)\psi\left(\frac{s-b}{a}\right)ds \quad (2)$$

where  $\psi$  is the mother wavelet;  $a$  is the scale factor that determines the characteristic frequency or wavelength; and  $b$  represents the shift of the wavelet over  $x(s)$ . In this study, the wavelet coefficient was scaled by  $1/a$ , which represents wavelet amplitudes while most other applications use  $1/\sqrt{a}$  to calculate the modulus or variance of the signals (Chao and Naito, 1995). The wavelet coefficients  $W$  were normalized to represent the amplitude of Fourier frequencies. The continuous wavelet transform (CWT) was used, with the Morlet wavelet as the mother function (Morlet et al., 1982). The shifted and scaled Morlet mother wavelet is defined as (Chao and Naito, 1995):

$$\psi_{l,a,b}(s) = \sqrt[4]{\pi}\sqrt{a}e^{\frac{i2\pi(s-b)}{a}}e^{-\frac{1}{2}\left(\frac{s-b}{a}\right)^2} \quad (3)$$

The Morlet wavelet is a sinusoid with wavelength/period  $a$  modulated by a Gaussian function, which provides robust results in analyses of climate and hydrological related records (e.g., Adamowski et al., 2009; Coulibaly, 2006; Torrence and Compo, 1998). A parameter  $l$  modifies wavelet transform bandwidth resolution either in favor of time or in favor of frequency, according to the uncertainty principle. The bandwidth (frequency) resolution for the wavelet transform varies with (Prokoph and Patterson, 2004)

$$\Delta a = \frac{a\sqrt{2}}{4\pi l} \quad (4)$$

and a location resolution

$$\Delta b = \frac{al}{\sqrt{2}} \quad (5)$$

The wavelet coefficients on the top and bottom of the data are subject to 'edge effects', because only half of the Morlet wavelet lies inside the data set. Missing data for the analysis window were replaced ('padded') by zeros. For long wavelengths (e.g., wavelength  $a$  covers more than half of the whole data series) the edge effect can stretch across the whole time series. Thus, the boundary of edge effects on the wavelet coefficients forms a wavelength dependent curve for 'weak edge affected' areas known as the 'cone of influence' (Torrence and Compo, 1998).

To transform a measured, and hence limited and discrete time series, the trapezoidal rule was used for the wavelet transform, which provides  $W^l(a, b)$ . The cross-wavelet spectrum of two series  $x(t)$  and  $y(t)$  is defined by:

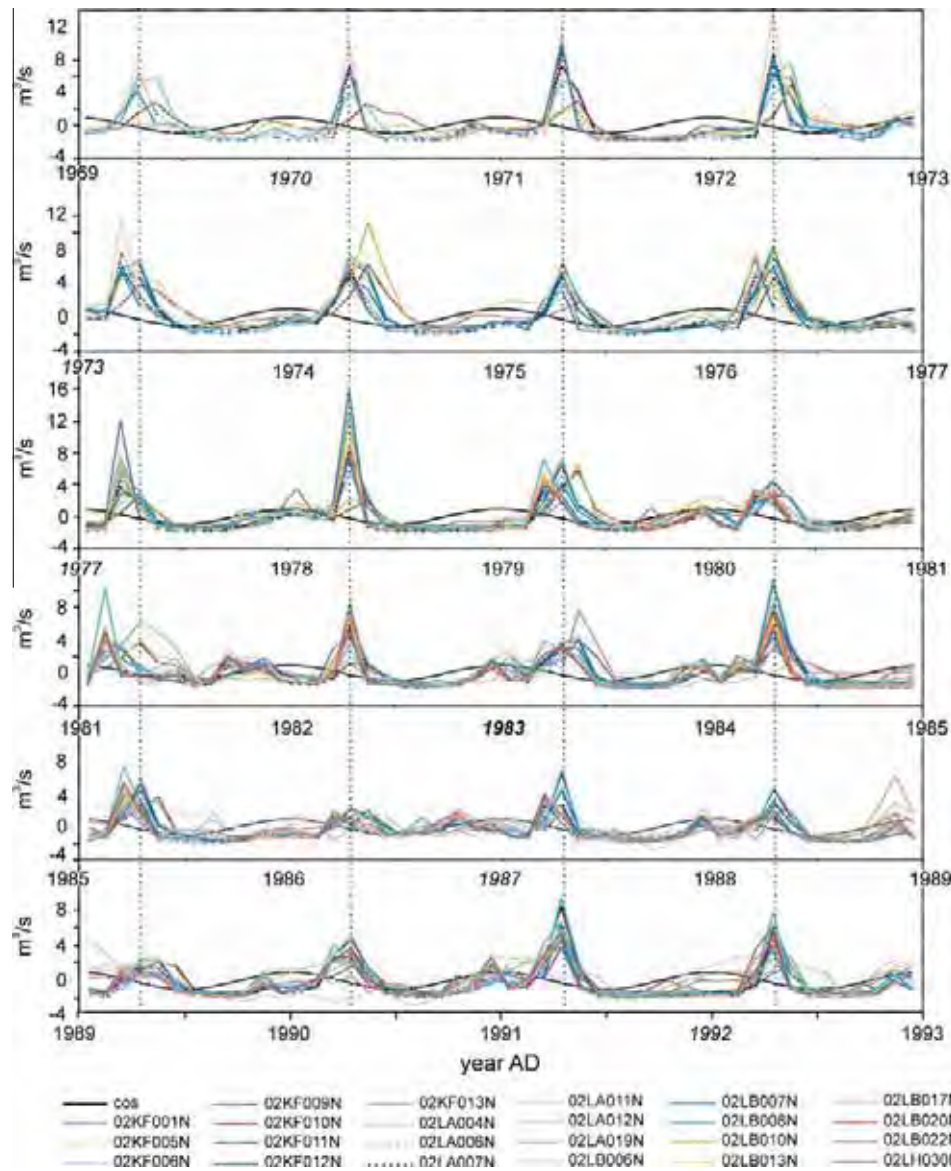
$$W_{xy}(a, b) = W_x(a, b)W_y^*(a, b) \quad (6)$$

where  $W_x(a, b)$  and  $W_y(a, b)$  are the CWT of  $x(t)$  and  $y(t)$  respectively, where  $*$  denotes the complex conjugate (Jury et al., 2002). In our application the modulus of the wavelet transform  $|W_{xy}(a, b)|$  represents the cross-amplitudes of  $x(t)$  and  $y(t)$ . The phase difference is defined by:

$$\Delta\Phi(b) = \tan^{-1} \frac{\int_{a_1}^{a_2} \text{im}(W_{xy}(a, b))da}{\int_{a_1}^{a_2} \text{Re}(W_{xy}(a, b))da} \quad (7)$$

with  $b$  corresponding to the time lag  $b$  (Jury et al., 2002).  $Im$  and  $Re$  indicate the imaginary and real part, respectively.





**Fig. 5.** Normalized monthly streamflow records for 23 stations (for ID's see Table 1) and cosine curve of amplitude 1 and wavelength of 12 months; El Niño years are marked in italic, very strong El Niño year (1983) marked in bold italics.

Wavelet analysis was carried out on the monthly streamflow records for a bandwidth of 0.8–1.2 years to extract annual signals ( $W_{1yr..b}$ ) only. As a result of the edge effect, the wavelet coefficients for a specific wavelength represent 94% of the Fourier amplitude that would be extracted from a stationary process using spectral analysis for a 12 month wavelength. To counteract the edge effect influence, wavelet analysis was also carried out on a standing cosine wave with amplitude 1, resulting in  $W_{cos..b}$ . Finally, the wavelet coefficient was corrected by:

$$W_{c1yr..b} = W_{1yr..b} / W_{cos..b} \quad (8)$$

In many applications the parameter  $l$  is chosen to be between  $l = 6$  and 10 (e.g., Prokoph and Patterson, 2004; Ware and Thomson, 2000) with sufficiently precise results in resolving time and frequency, respectively. In this study, wavelet analysis was applied to hydrological records with many gaps. Even small gaps may lead to a reduction of amplitude and 'smearing' effects on the onset and offset of gaps by using  $l \geq 6$ . Thus,  $l = 2$  was chosen for superior amplitude recognition in time (location), even if frequency

resolution is diminished compared to  $l = 6$ . The reduced frequency resolution (see Eq. (3)) due to using  $l = 2$  is of minor concern for this study, because the narrow bandwidth (0.8–1.2 years) analyzed by wavelet transform sufficiently represents the dominant annual signal in the streamflow variability (see Fig. 5). Furthermore, cross-wavelet coefficients and phase shifts were calculated to a amplitude = 1 and wavelength = 12 month signal to (a) emphasize the annual wavelength and (b) to replace phase by phase shift, which is easier to use to calculate phase change trends over time. The amplitude and wavelength pattern relevant to this study that can be extracted by cross-wavelet transform is shown via a conceptual model that is composed of a synthetic annual signal with time gaps of different lengths (Fig. 7).

### 3.3. GIS analysis

ESRI ArcView 3.2 with spatial analyst module was used for spatial–temporal analysis of the 9 meteorological records. The following protocol was followed:

**Table 1**  
Meteorological and streamflow stations.

Station name	North	West	Time period		
			Temperature	Precipitation	
Meteorological station					
Arnprior	45°25'	76°22'	7/1964–9/1999	1/1965–9/1999	
Maniwaki UA	46°18'	76°00'	8/1953–5/1993	8/1953–5/1993	
Wakefield	45°36'	75°54'	1/1963–3/1993	1/1963–3/1993	
Ottawa CDA	45°23'	75°43'	11/1889–8/2002	11/1889–8/2002	
Morrisburg	44°55'	75°11'	3/1913–8/2002	3/1913–8/2002	
Cornwall	45°01'	74°45'	11/1950–8/2002	11/1950–8/2002	
Glen Gordon	45°10'	74°32'	1/1969–7/1999	1/1967–7/1999	
Dalhousie mills	45°19'	74°28'	5/1968–8/2002	5/1968–8/2002	
LesCedres	45°18'	74°03'	3/1913–8/2002	3/1913–8/2002	
Streamflow station	North	West	Map code	Station ID	Drainage area (km <sup>2</sup> )
Clyde River at Gordon Rapids	45°08'	76°38'	1	02KF013	280
Clyde River near Lanark	45°03'	76°24'	2	02KF010	614
Mississippi River at Fergusons Falls	45°03'	76°17'	3	02KF001	2620
Indian River near Blakeney	45°15'	76°16'	4	02KF012	203
Ottawa River at Chats Falls	45°28'	76°14'	5	02KF009	89,600
Carp River near Kinburn	45°25'	76°12'	6	02KF011	269
Mississippi River at Appleton	45°10'	76°07'	7	02KF006	2900
Rideau River below Merrickville	44°57'	75°49'	8	02LA011	1920
Ottawa River at Britannia	45°23'	75°48'	9	02KF005	90,900
Jock River near Richmond	45°15'	75°47'	10	02LA007	559
Gatineau (Riviere) a la Centrale des Rapides Farmers	45°30'	75°46'	11	02LH030	23,600
Outaouais (Riviere) des Centrale de Hull 2	45°25'	75°44'	12	02LA019	91,300
Rideau River at Ottawa	45°23'	75°42'	13	02LA004	3830
Rideau River below Manotick	45°15'	75°42'	14	02LA012	3120
Kemptville Creek near Kemptville	44°60'	75°40'	15	02LA006	409
South Nation River at Spencerville	44°50'	75°33'	16	02LB007	246
North Branch South Nation River near Heckston	44°59'	75°31'	17	02LB017	69.2
South Castor River at Kenmore	45°14'	75°25'	18	02LB020	189
Ottawa River at Cumberland	45°31'	75°24'	19	02LB010	92,000
Castor River at Russell	45°16'	75°21'	20	02LB006	433
Bear Brook near Bourget	45°26'	75°09'	21	02LB008	440
Payne River near Berwick	45°12'	75°06'	22	02LB022	152
South Nation River at Casselman	45°19'	75°05'	23	02LB013	2410

### 3.3.1. Data pre-processing

- Four year average temperature and precipitation were calculated for 6 intervals from 1/1969 to 12/1992 for the 9 meteorological stations and average annual amplitudes and phase shifts for up to 23 streamflow records per 4 year interval. For easier interpretation, the phase shift of the annual streamflow cycle was transformed from radian  $\phi_r$  into days  $\phi_d$  by  $\phi_d = \phi_r (365.25/2\pi)$ , applying an average year of 365.25 days.
- Linear regression equations  $y(t) = a + b(t)$  with  $a$  intercept at  $t = 0$ , and  $b$  gradient were calculated by Least-square-error estimate for each of the 24-year temperature and precipitation records for each of the 9 meteorological stations. In

addition linear regression gradients were calculated in the 24-year average annual amplitudes and phase shifts that were derived from wavelet analysis for the 9 streamflow records that have no record gaps.

### 3.3.2. ArcView3.2 data processing

The SPATIAL ANALYST module was used for calculating the gradient map for each 4 year interval in temperature and precipitation and their gradients by using an 'inverse squared distance' interpolation method applying a 150 km range of influence for this interpolation, and using the location range as analysis areas.

The hydrological parameters, which includes 4 year averaged annual streamflow amplitude and annual variation (i.e., phase shift) of flood onset, were plotted at their location and color coded according to relative amplitude and phase shift.

## 4. Results

### 4.1. Temporal variability in annual streamflow

The normalized streamflow records show good accordance between all records, particularly in terms of their flood peaks, except for the late 1980s and during 1983 when an El Nino influence was strong in Eastern Canada (Fig. 5). Besides the spring flood peak, most streamflow records also exhibit autumn peaks, and especially since ~1980 (Fig. 5). The wavelet coefficients (amplitudes) of the annual streamflow signals show strong year-to-year variability from 1969 to 1985, and are relatively stable after 1986 (Fig. 8). The year 1986, which is set as a calibration year, has the weakest annual amplitude in the entire 24 year time interval, while the year 1984 has the strongest annual amplitude as can be seen in Fig. 5. Most stations show similar variability, but the annual amplitudes of stations 02LB010N (Ottawa River at Cumberland) and 02KF009N (Ottawa River at Chats Falls) are often different, probably due to water flow regulations on the Ottawa River.

In general, the cross-wavelet analysis derived phase shifts show good accordance between all stations, except for 02LB010N (Ottawa River at Cumberland) and 02KF009N (Ottawa River at Chats Falls), that are slightly off to all other stations for most of the years (Fig. 9). The years 1982 and 1987 show a strong deviation from the normal phase shift (~1–2 radians) due to some unusually large autumn floods that suppressed the dominance of the spring floods in these years.

The four year average of phase shifts for each hydrological station show that the phase shift generally increased. The phase shifts peak in the early 1980s, indicating earlier onsets of spring floods and/or generally higher streamflows occurring in the earlier part of the year (Fig. 10A: low radians-late annual floods, high radians-early annual floods). The increasing trend in phase shift changes occurs in all stations, whereas the four year peak interval varied (Fig. 10A). Similarly, the four year average streamflow amplitude variability demonstrates common peaks in the mid 1970s, and lows in the mid 1980s, with strong local differences (Fig. 10B). For example, the annual streamflow amplitude varied little for station 02KF005 (Ottawa River at Britannia), probably due to dams and other streamflow regulations along the Ottawa River, whereas stations in the more rural areas on pristine rivers such as 02LB006 (Castor River at Russell) varied in relative amplitude between 2.5 and 1.5 compared to 1986, and are generally higher than the four year amplitude average from all stations (Fig. 10B).

### 4.2. Spatial-temporal meteorological and annual streamflow variability

Spatial analysis of the four year average annual streamflow amplitude and air temperature in eastern Ontario/southwestern



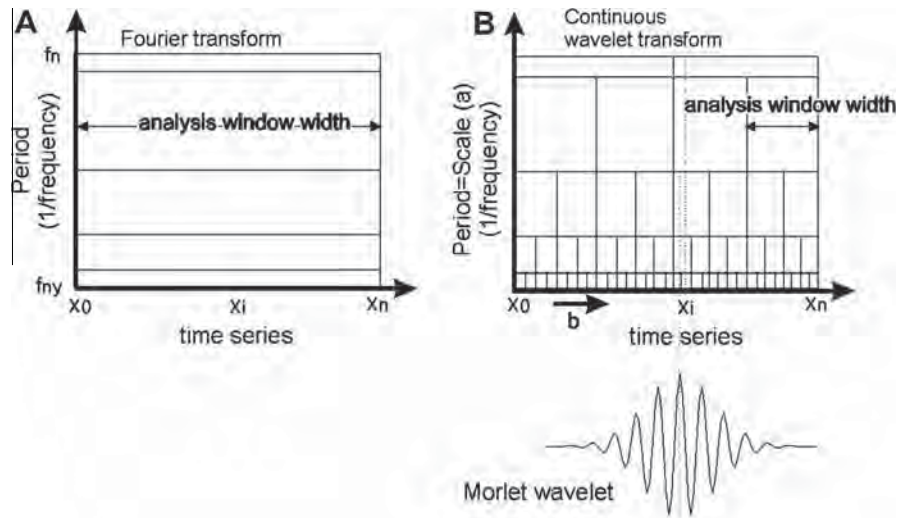


Fig. 6. Morlet transform scaling: Top: Analysis window (AW) in scale-location space. Bottom: Morlet wavelet centered at location  $b_1$ .

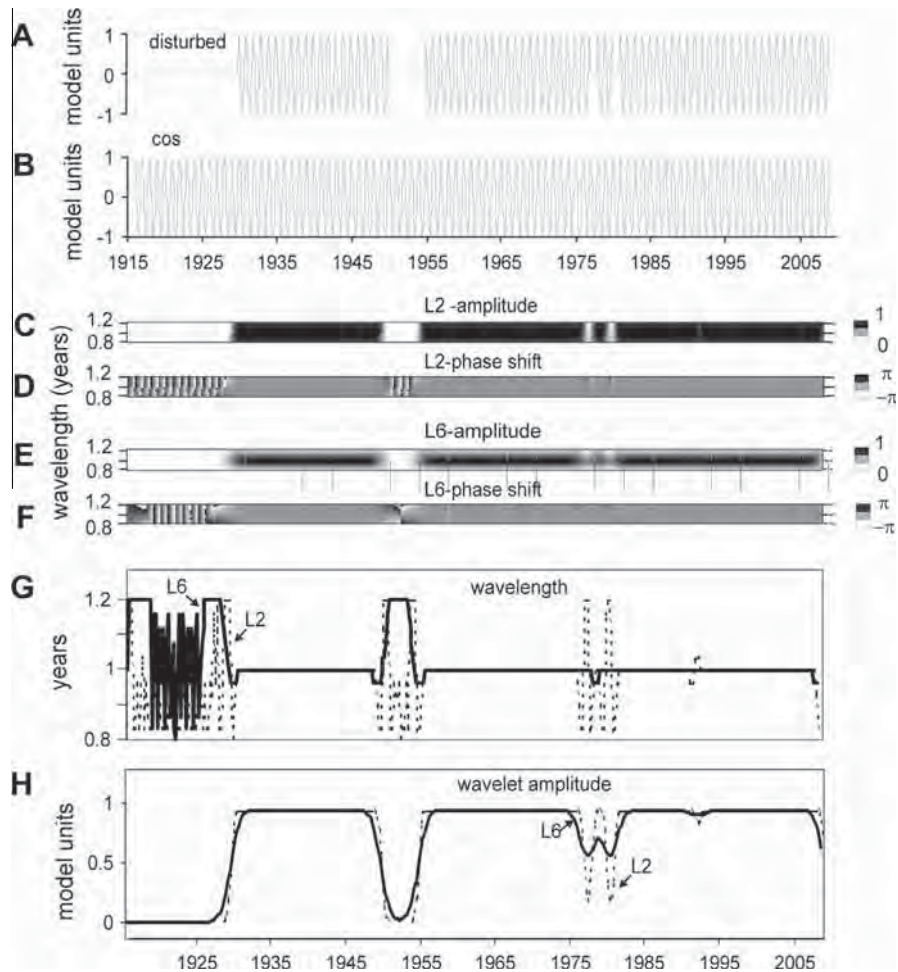


Fig. 7. Cross-wavelet analysis of conceptual model for model years 1916–2010 and centered at waveband 0.8–1.2 years: (A) conceptual streamflow model with 1 year wavelength, amplitude = 1 model unit and missing (nulled) record of different duration, (B) standing cosine wave of 1 year wavelength, amplitude = 1 model unit, (C) cross-wavelet amplitude using AW scaling factor  $L = 2$ , (D) phase shift between conceptual model and standing cosine wave using AW scaling factor  $L = 2$ , (E) cross-wavelet amplitude using AW scaling factor  $L = 6$ , (F) phase shift between conceptual model and standing cosine wave using AW scaling factor  $L = 6$ , (G) wavelength with strongest local (cross-wavelet) amplitude for  $L = 2$  and  $L = 6$ , (H) locally strongest (cross-wavelet) amplitude, corrected for edge effects, for  $L = 2$  and  $L = 6$ .

Quebec is shown for a relatively cool interval (1977–1980) and a relatively warm interval (1985–1988) (Fig. 11). In both intervals,

a strong regional NW–SE temperature gradient existed, which was already punctuated by the urban heat islands of Ottawa,

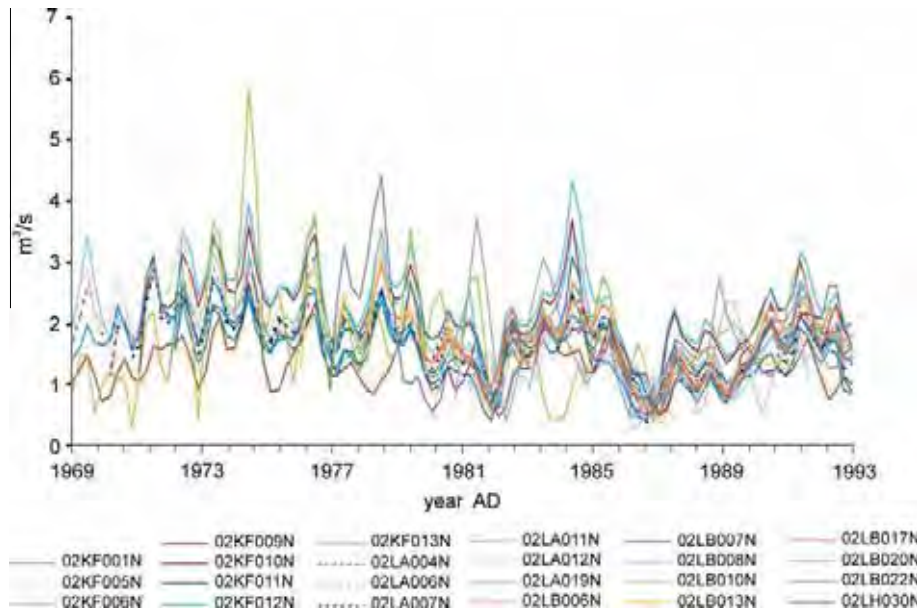


Fig. 8. Edge-effect corrected amplitudes of normalized streamflow for 23 stations (for ID's see Table 1) from 1969 to 1992.

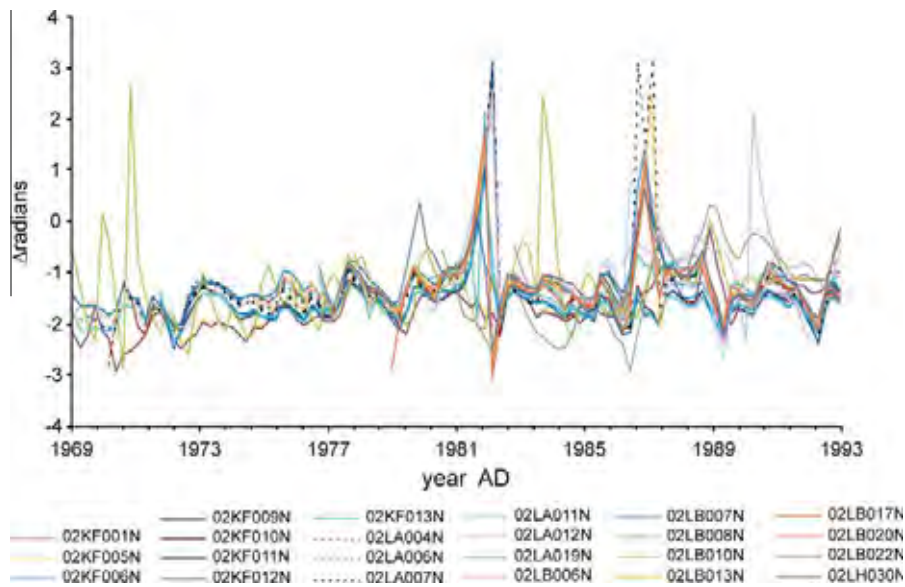


Fig. 9. Phase shift of normalized streamflow to 12-month cosine wave for 23 stations (for ID's see Table 1) from 1969 to 1992.

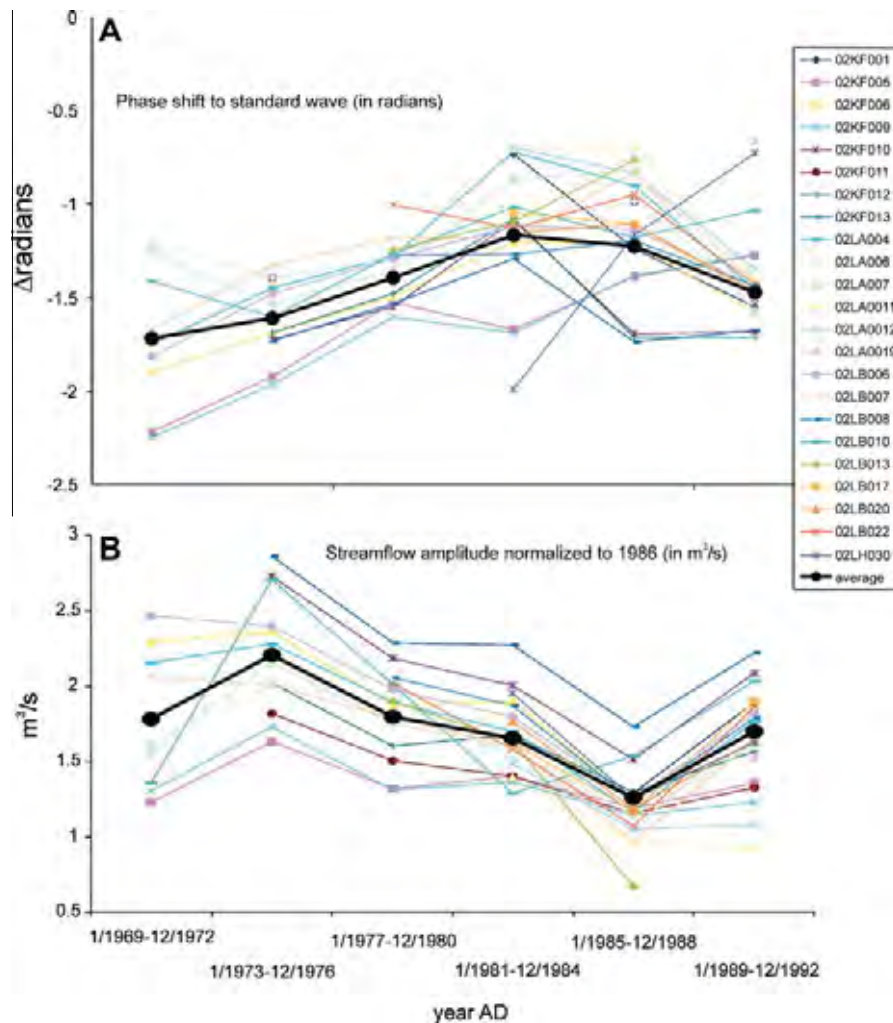
Montreal and Cornwall during the relatively cool period; they were already >0.3 °C warmer than their surroundings (Fig. 11). The dominance of the heat island decreased through the 1970s to the 1990s, because of the warming of the region of southeastern Ontario by ~0.5 °C (Fig. 11). In the transition from “cool” to “warm” the relative annual streamflow amplitude generally decreased, with higher amplitudes remaining in the cooler western parts (Fig. 11).

The mean precipitation maps for the wettest 4-year period (1969–1972) and driest period (1985–1988) show that (1) the precipitation distribution was spatially more even during the wet than dry periods in eastern Ontario/southwestern Quebec (Fig. 12), (2) an increasing precipitation trend exists from east to west, (3) reduced precipitation occurred in the heat islands of Ottawa and Montreal, and (4) a positive precipitation anomaly existed northwest of Ottawa due to a ~30 km long mountain range with topography up to 900 m above sea level. The annual streamflow cycle

peaked up to 50 days later in the low precipitation western regions compared to the high precipitation area in the southeast of the region (Fig. 12).

In the study area, a positive warming trend is common for all locations but most pronounced and significant in the cities. In contrast, the precipitation trends vary from positive to negative and are generally less significant (Table 2).

All nine hydrological stations with a complete 24 year record show decreasing annual amplitudes in streamflow (i.e., less pronounced spring flooding or generally less seasonality in the streamflow). The average decrease in amplitude by 0.1 compared to a normalized standard deviation of 1 for the annual flow of 1986 (see Eq. (1)) indicates a  $0.1/1 * 100\% = 10\%$  reduction of annual streamflow amplitude compared to the individual 1986 streamflow amplitude. This drop is >95% significant and stronger (~20%) in the smaller, rural streams such as the Castor River near



**Fig. 10.** Normalized streamflow pattern for 23 stations (for ID's see Table 1) from 1969 to 1992: (A) 4 year averaged annual normalized streamflow amplitudes; (B) 4 year averaged phase shift for annual normalized streamflow to 12-month standard wave; Bold black line indicates averages from 9 stations with complete 1969–1992 records.

Russell and comparably insignificant in the broader, more regulated streams, such as the Ottawa River near Britannia (~2%).

Phase-lag records extracted by cross-wavelet analysis show that there is no significant phase shift towards earlier spring floods or a redistribution of total annual water flow towards the beginning of the year in the rural areas, with a range from 0 to ~1.2 days/year, which results in a total of <1 month over the entire 24 year survey period. In contrast, the Ottawa River stations experienced >95% significant early onset of spring flow by ~2.5 days/year or ~50 days, which is almost 2 months over the entire survey period.

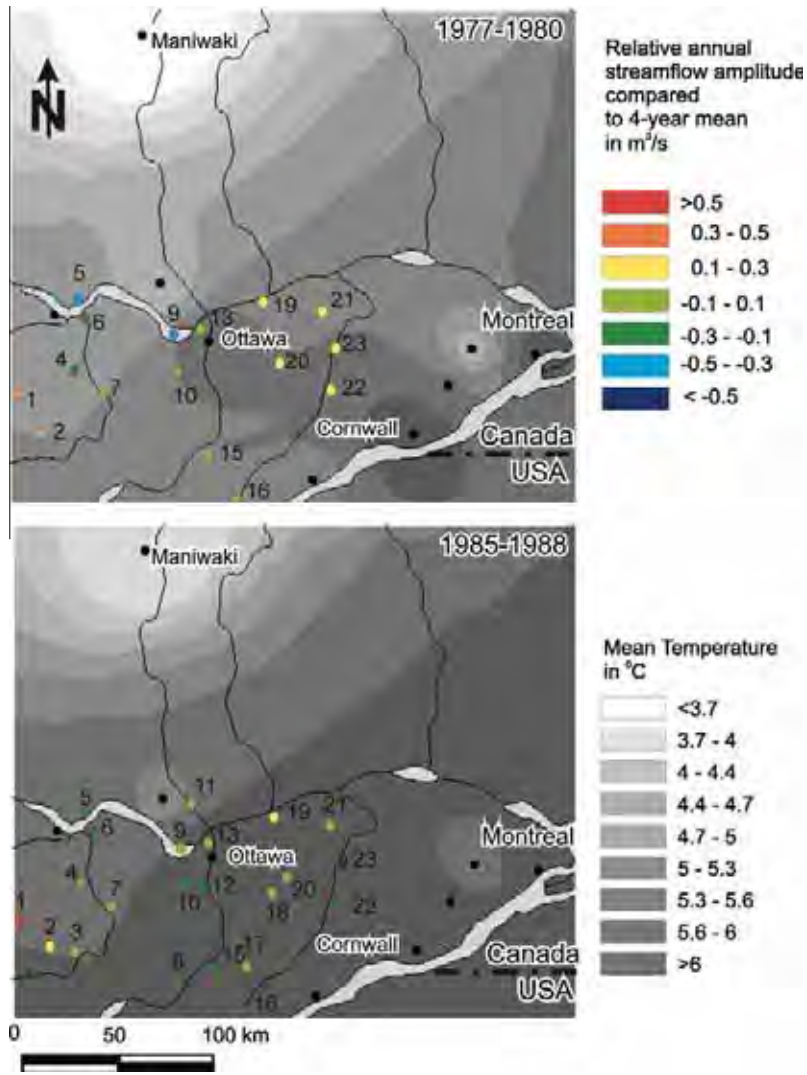
Fig. 13 shows that the 1969–1992 gradients in air temperature increased from northwest to southeast, whereas the precipitation gradients decreased in the same direction, indicating that the South easternmost region of Ontario warmed up and received less precipitation. Furthermore, the annual streamflow amplitude increased slightly over the entire 1969–1992 period in the north-western part, and decreased towards the southeast (Fig. 13).

The spatial pattern of change of annual streamflow cycle offset follows those northwest–southeast trends in air temperature and precipitation change. The average offset is 1.09 days/years. A relatively higher tendency of earlier flooding and/or generally higher streamflow in the beginning of the year occurs along the Ottawa River and in the western region, in accordance with similarly increasing precipitation in these parts of the region.

## 5. Discussion

This study used cross-wavelet analysis (XWT) with the Morlet wavelet as the mother wavelet to extract the cross-wavelet spectrum representing (annual) amplitudes of streamflow, as well as to determine the phase shift (timing of annual flow maximum) through a 24-year period using monthly mean streamflows. Based on Heisenberg's uncertainty principle, time (location) and frequency cannot be perfectly resolved at the same time. The result of a perfect frequency resolution as potentially provided by the Fourier transform would lose all time resolution (e.g., Rioul and Vetterli, 1996), which is essential for this study. In this study, time resolution was set to be very precise ( $\pm 0.5$  years), while resolution in the annual frequency band was set low. Maraun and Holschneider (2005) highlighted the problems arising from XWT between a signal and noise resulting in a less persistent, but still relatively strong, cross-wavelet spectrum in the frequency of the record with the real signal, which after testing several time frequency resolutions still appeared statistically significant. In this study, it is already visually evident from the normalized streamflow records (Fig. 5) that the annual signal, dominated by the annual spring flood, appears clearly above the background. This allows the XWT to determine and extract the annual signal relatively precisely in frequency. In addition, the high time precision allowed for the automatic delineation of time intervals without





**Fig. 11.** GIS maps (with main watersheds) of 4 year mean air temperature (gray contours) overlaid by color coded 4 year mean difference of streamflow amplitudes to 4 year mean (see Fig. 10). For locations see Table 1 and Fig. 1. Top: 4-year interval with lowest average temperature in region from 1969 to 1992. Bottom: 4-year interval with highest average temperature in region from 1969 to 1992.

records for removal from further treatment. The monthly records used in this study had the advantage that (1) the amplitude of annual springflood is smoothed out thus making the Morlet wavelet a good choice as the mother wavelet, and (2) they made the timing of the springflood onset less ambiguous due to the combination of the pre-spring flood pulse with the main spring flood pulse as found in western North America (e.g., Steward et al., 2005). Such a pre-spring flood pulse pattern is also evident in the streamflows in the Ottawa area (e.g., Adamowski, 2008).

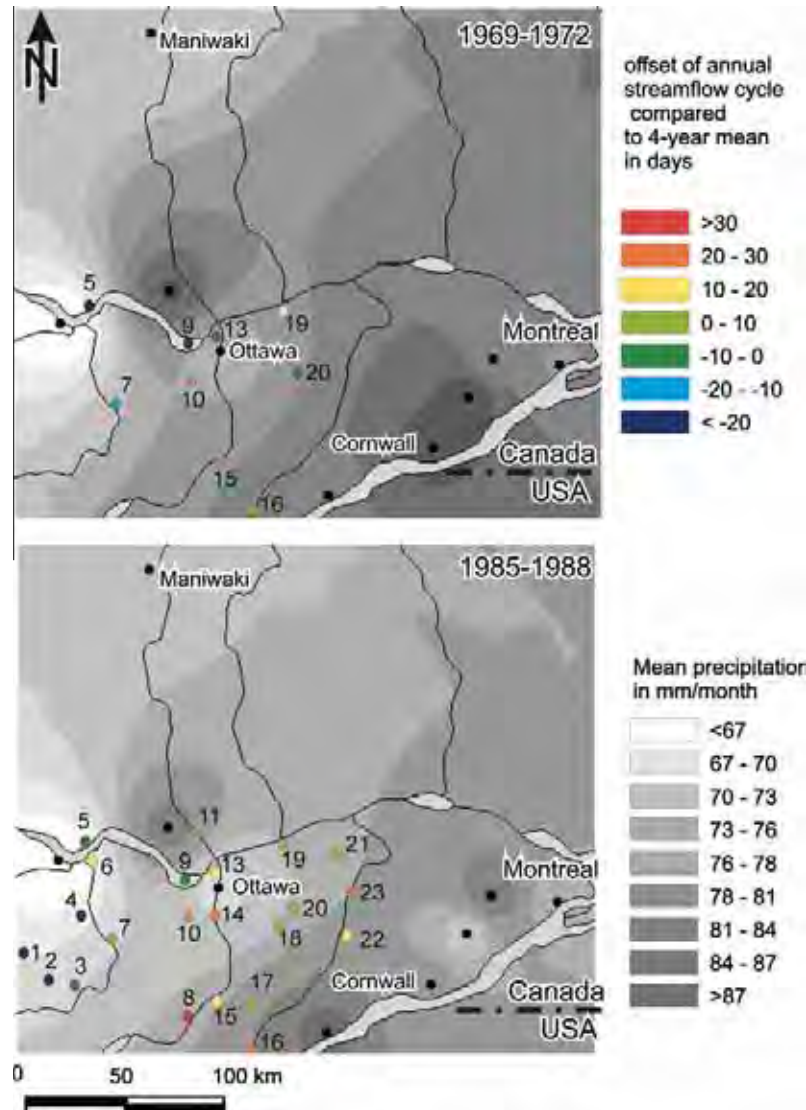
GIS analysis was combined with spatial trend analysis for meteorological data using inverse square weightings and color coding of trends and pattern of sources for streamflow stations. In this study, the temperature and total precipitation trends emerge from the data from the nine meteorological stations, and reflect the relatively gradual change in meteorological conditions across the region. The NW–SE trends in temperature and precipitation (see Figs. 11 and 12) are evident in several station records. Warm and dry anomalies around major cities such as Ottawa and Montreal highlight the well pronounced urban heat islands around these cities that have been identified already (e.g., Prokoph and Patterson, 2004). Over the 24 year interval explored in this study, the warming trend was enhanced towards the southeast, while precipitation

was reduced in this direction (Fig. 13). This confirms the results for the USA as shown by Krakauer and Fung (2008), who found that precipitation in NE USA increased in the 1960s but afterwards did not show any significant trend. In particular, an increasing precipitation toward the east was identified.

The results of this study also show that in the same time interval the annual flow amplitude, which roughly represents the range between annual minimum and maximum flow, did not change significantly, and may be multi-decadal cyclic (Fig. 10). This confirms a similar trend fluctuation for the USA (Lins and Slack, 1999; McCabe and Wolock, 2002), and for southeastern Canada (Burn et al., 2010). The results of this study show that not only is the annual streamflow amplitude trend variable, and in general insignificant in eastern Ontario/western Quebec, but also that the larger rivers (i.e., Ottawa River and Rideau River) have relatively less annual amplitude fluctuations than the smaller waterways. This may be due to waterway regulations on the larger rivers, while the smaller rivers are more pristine.

The annual streamflow maximum occurred on average ~25 days earlier in 1969 than in 1992 at a rate of change of ~1.09 days/year (Fig. 10), confirming similar findings for other regions in Canada (Zhang et al., 2001) and the USA (e.g., Steward





**Fig. 12.** GIS maps (with main watersheds) of 4 year mean monthly total precipitation (gray contours) overlaid by color coded 4 year mean difference of annual streamflow phase shift to 4 year mean (see Fig. 10). For locations see Table 1 and Fig. 1. Top: 4-year interval with highest average precipitation in region from 1969 to 1992. Bottom: 4-year interval with lowest average precipitation in region from 1969 to 1992.

**Table 2**  
Meteorological trends.

Location	North	East	$\Delta$ temperature		$\Delta$ annual average monthly precipitation	
			Celsius/year	<i>r</i>	mm/year	<i>r</i>
Arnprior	45°25'	76°22'	0.0379	<b>0.88</b>	0.1079	0.35
Cornwell	45°01'	74°45'	0.0577	<b>0.92</b>	-0.1232	-0.32
Dalhousie	45°19'	74°28'	0.0474	0.76	-0.0619	-0.74
Glen Gordon	45°10'	74°32'	0.0389	0.72	0.1686	0.48
LesCedres	45°18'	74°03'	0.0337	<b>0.83</b>	-0.3	-0.62
Maniwaki UA	46°18'	76°00'	0.0244	0.60	-0.6888	<b>-0.90</b>
Morrisburg	44°55'	75°11'	0.0464	0.72	0.2605	0.22
Ottawa	45°23'	75°43'	0.0428	<b>0.88</b>	-0.2146	-0.37
Wakefield	45°36'	75°54'	0.0385	0.69	0.0748	0.20

*r*: Linear correlation coefficient.

Bold: correlation coefficient *r* > 95% significant for 6 samples.

et al., 2005). In addition, this study found that this trend towards earlier annual maximum flows increased strongly in the north-western part of the study area, with less warming but increasing precipitation. This suggests that increases in precipitation occurring earlier in the year lead to maximum flows earlier in the year.

## 6. Conclusions

This study applied time series analysis (CWT and XWT), and Geographic Information System (GIS) techniques, to evaluate the potential influence of long term climate change on annual water

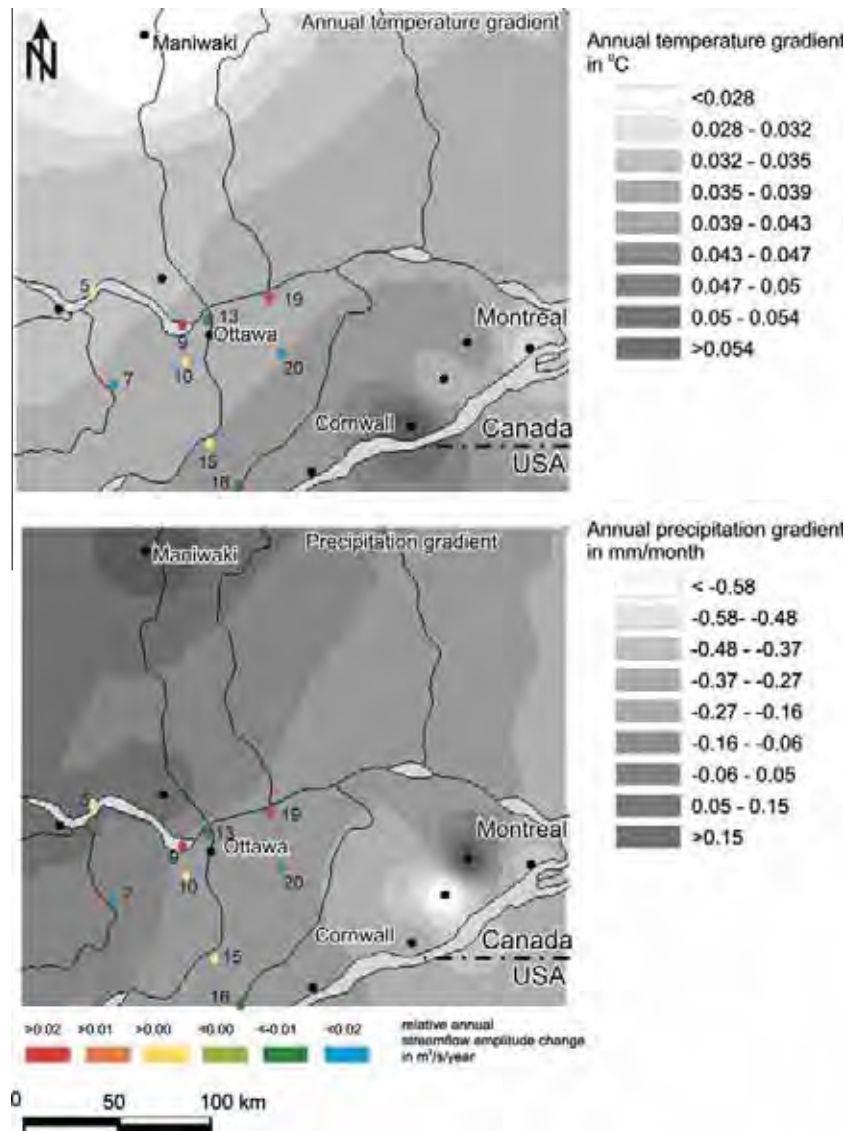


Fig. 13. GIS maps (with main watersheds) of 1969–1992 trends (top) of mean air temperature (gray contours) and (bottom) of mean monthly total precipitation (gray contours) overlaid by color coded trends of mean difference of annual streamflow amplitudes relative to mean. For locations see Table 1 and Fig. 1.

distributions in eastern Canada for the 24 year period from 1969 to 1992. The results show that over the 24 year period the southwestern part along the US Border experienced an increase in temperature of up to 0.05 °C/year, and decreased precipitation by ~0.5 mm/month, while the more rural northwest region experienced insignificant warming (<0.02 °C/year) and became only <0.2 mm/year wetter. In the same time interval, the annual streamflow shifted on average to a ~20 days earlier flood onset following a similar northwest-southeast trend as the meteorological records. In the northwest, along the Ottawa River, floods occurred in 1992 ~50 days earlier than in 1969, while only <10 days earlier than in the southeastern streams.

The streamflows of the main waterways in the region (Ottawa and Rideau River) generally have relatively lower annual amplitudes than the smaller and more pristine waterways that experienced more seasonality. The annual streamflow amplitude is, in general, decreasing but shows only significant spatial-temporal trends through the 24 year interval in small watersheds with low streamflow volume, similar to findings in the USA and other parts of Canada (e.g., Burn et al., 2008, 2010).

The results suggest the need for water resource managers to adjust to a higher probability of earlier flood onsets, with the highest water flow volumes generally occurring earlier in the year; this is particularly significant in the northwestern regions and in streams with large drainage areas with a more distinct continental climate. With the southeastern regions experiencing more maritime influences, a wetter climate provides higher streamflows later in the year (summer/fall). As such, a shift in the general annual water volume distribution will be less pronounced.

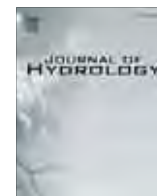
**Acknowledgements**

The research in the article was funded by a NSERC Discovery Grant, and a CFI grant, held by Jan Adamowski. We thank P. Larson from the Rideau River Valley Conservation authority for providing streamflow data.

**References**

Adamowski, J., 2008. River flow forecasting using wavelet and cross-wavelet transform models. *Hydrological Processes* 22, 4877–4891.

- Adamowski, K., Bocci, C., 2001. Geostatistical regional trend detection in river flow data. *Hydrological Processes* 15, 3331–3341.
- Adamowski, K., Prokoph, A., Adamowski, J., 2009. Development of a new method of wavelet aided trend detection and estimation. *Hydrological Processes* 23, 2686–2696.
- Adamowski, J., Prokoph, A., Adamowski, K., 2011. Spatial temporal changes in streamflow patterns in eastern Ontario and southwestern Quebec, Canada and their relation to precipitation changes. *The International Journal of Climate Change: Impacts and Responses* 3, 155–170.
- Bonsal, B., Zhang, X., Vincent, L., Hogg, W., 2001. Characteristics of daily and extreme temperatures over Canada. *Journal of Climate* 14, 1959–1976.
- Burn, D.H., Hag Elnur, M., 2002. Detection of hydrologic trend and variability. *Journal of Hydrology* 255, 107–122.
- Burn, D.H., Fan, L., Bell, G., 2008. Identification and quantification of streamflow trends on the Canadian Prairies. *Hydrological Sciences Journal* 53, 538–549.
- Burn, D.H., Sharif, M., Zhang, K., 2010. Detection of trends in hydrological extremes for Canadian watersheds. *Hydrological Processes* 24, 1781–1790.
- Chao, B.F., Naito, I., 1995. Wavelet analysis provides a new tool for studying Earth's rotation. *EOS* 76, 164–165.
- Coulibaly, P., 2006. Spatial and temporal variability of Canadian seasonal precipitation. *Advances in Water Resources* 29, 1846–1865.
- Coulibaly, P., Burn, D.H., 2005. Spatial and temporal variability of Canadian seasonal streamflows. *Journal of Climate* 18, 191–210.
- Couture, R., Gauvreau, D., Belanger R.J. 2006. Development of a 3-D geological model towards natural hazards mitigation, St. Lawrence River Valley, Eastern Canada. *IAEG 2006 Paper 788*: 1–13.
- Cunderlik, J.M., Burn, D.H., 2002. Local and regional trends in monthly maximum flows in southern British Columbia. *Canadian Water Resources Journal* 27, 191–212.
- Davis, J.C., 1986. *Statistics and Data Analysis in Geology*. Wiley, New York, p. 646.
- Ehsanzadeh, E., Adamowski, K., 2010. Trend in timing of low stream flows in Canada: impact of autocorrelation and long-term persistence. *Hydrological Processes* 24, 970–980.
- IPCC, 2007. In: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. (Eds.), *Climate Change 2007: Working Group I Report: the Physical Science Basis*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Jury, M.R., Enfield, D.B., M elice, J., 2002. Tropical monsoons around Africa: stability of El Ni no–Southern Oscillation associations and links with continental climate. *Journal of Geophysical Research* 107, 3151–3167.
- Karl, T.R., Diaz, H.F., Kukla, G., 1988. Urbanization: its detection and effect in the United States climate record. *Journal of Climate* 1, 1099–1123.
- Ko, C., Cheng, Q., 2004. GIS spatial modeling of river flow and precipitation in the Oak Ridges Moraine area, Ontario. *Computers and Geosciences* 30, 379–389.
- Krakauer, N.Y., Fung, I., 2008. Is streamflow increasing? Trends in the coterminous United States. *Hydrological Earth System Sciences Discussion* 5, 785–810.
- Lins, H.F., Slack, J., 1999. Streamflow trends in the United States. *Geophysical Research Letters* 26, 227–230.
- Maraun, D., Holschneider, M., 2005. *Wavelet Spectral Analysis: Developments and Examples*. University of Potsdam Publication, Germany, p. 54.
- McCabe, G.J., Wolock, D.M., 2002. A step increase in streamflow in the conterminous United States. *Geophysical Research Letters* 29, 788–791.
- Morlet, J., Arehs, G., Fourgeau, I., Giard, D., 1982. Wave propagation and sampling theory. *Geophysics* 47, 203–236.
- Pilon, P.J., Yue, S., 2002. Detecting climate-related trends in streamflow data. *Water Science Technologies* 8, 89–104.
- Prokoph, A., Patterson, R.T., 2004. Application of wavelet and regression analysis in assessing temporal and geographic climate variability: eastern Ontario, Canada as a case study. *Atmosphere-Ocean* 42, 201–212.
- Rioul, O., Vetterli, M., 1996. Wavelets and signal processing. *IEEE Special Magazine*, 14–38.
- Steward, I.T., Cayan, D.R., Dettinger, M.D., 2005. Changes toward earlier streamflow timing across Western North America. *Journal of Climate* 18, 1136–1155.
- Telmer, K., Veizer, J., 1996. Natural and human signals in the Ottawa River, Canada, and its tributaries: GIS and statistical analysis. *Anais VIII Simp osio Brasileiro de Sensoriamento Remoto, Salvador, Brasil, 14–19 abril 1996*, INPE, 407–408.
- Torrence, C., Compo, G.P., 1998. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society* 79, 61–78.
- Ware, D.M., Thomson, R.E., 2000. Interannual to multidecadal timescale climate variations in the Northeast Pacific. *Journal of Climate* 13, 3209–3220.
- Zhang, X., Vincent, L.A., Hogg, W.D., Niitsoo, A., 2000. Temperature and precipitation trends in Canada during the 20th century. *Atmosphere-Ocean* 38, 395–429.
- Zhang, X., Harvey, K.D., Hogg, W.D., Yuzyk, T., 2001. Trends in Canadian streamflow. *Water Resources Research* 37, 987–998.



# Multiscale streamflow forecasting using a new Bayesian Model Average based ensemble multi-wavelet Volterra nonlinear method



Maheswaran Rathinasamy<sup>a</sup>, Jan Adamowski<sup>b,\*</sup>, Rakesh Khosa<sup>a</sup>

<sup>a</sup> Department of Civil Engineering, Indian Institute of Technology, New Delhi, India

<sup>b</sup> Department of Bioresource Engineering, McGill University, 2111 Lakeshore, Ste Anne de Bellevue, Quebec H9X3V9, Canada

## ARTICLE INFO

### Article history:

Received 10 May 2013

Received in revised form 12 September 2013

Accepted 19 September 2013

Available online 5 October 2013

This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Sheng Yue, Associate Editor

### Keywords:

Multiscale streamflow forecasting

Wavelet based nonlinear models

Ensemble forecasting

Bayesian Model Averaging

## SUMMARY

Over the last five years, wavelet transform based models have begun to be explored for hydrologic forecasting applications. In general, a particular wavelet transform (and a particular set of levels of decomposition) is selected as the 'optimal' wavelet transform to be used for forecasting purposes. However, different wavelets have different strengths in capturing the different characteristics of particular hydrological processes. Therefore, relying on a single model based on a single wavelet often leads to predictions that capture some phenomena at the expenses of others. Ensemble approaches based on the use of multiple different wavelets, in conjunction with a multi model setup, could potentially improve model performances and also allow for uncertainty estimation. In this study, a new multi-wavelet based ensemble method was developed for the wavelet Volterra coupled model. Different wavelets, levels of decomposition, and model setups are used in this new approach to generate an ensemble of forecasts. These ensembles are combined using Bayesian Model Averaging (BMA) to develop more skilful and reliable forecasts. The new BMA based ensemble multi-wavelet Volterra approach was applied for forecasting stream flow at different scales (daily, weekly and monthly) observed at two stations in the USA. The results of this study reveal that the proposed BMA based ensemble multi-wavelet Volterra nonlinear model outperforms the single best wavelet Volterra model, as well as the mean averaged ensemble wavelet Volterra model.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

Wavelet transforms (WT) have begun to be used in several hydrological applications, such as trend analysis (e.g., Kallache et al., 2005; Partal and Küçük, 2006; Adamowski et al., 2009; Nalley et al., 2012; Sang et al., 2012) forecasting (e.g. Wang and Ding, 2003; Nourani et al., 2008; Kisi, 2009; Partal, 2009; Tiwari and Chatterjee, 2010; Adamowski and Sun, 2010; Adamowski and Fung Chan, 2012), time frequency characterisation (e.g., Labat, 2005; Neupauer et al., 2006; and Chen and Xie, 2007), multiscale correlation (e.g., Labat, 2005; and Sang et al., 2010c), and minimizing model parameters (Chou, 2007). In particular, the WT has been shown to be a very useful tool for hydrologic forecasting purposes. Recent publications (e.g., Adamowski and Karapataki, 2010; Kisi, 2011; and Maheswaran and Khosa, 2011c) have shown that the coupling of traditional data driven (or machine learning) models with the wavelet transform significantly increase the accuracy of these models. This has led to the development of coupling nonlinear models (e.g., neural networks, Support Vector Machines), as well as linear regression techniques (e.g., multiple linear

regression), with the wavelet coefficients resulting from filter bank decomposition. The majority of studies have found that the wavelet neural network approach provides the most accurate hydrological forecasts. More recently, Maheswaran and Khosa (2011c) developed a new wavelet Volterra coupled model that overcomes the limitations of the wavelet neural network model approach and provides more accurate results than the wavelet neural network approach.

Although significant progress has been made in the use of WTs for hydrologic and other time series forecasting, the choice of appropriate settings for the filter bank is still an open problem. Such settings involve the selection of a wavelet family, such as Daubechies (db), Symlet (sym) or Spline, an order (i.e. wavelet length) within that family, and the number of decomposition levels to be employed. Some authors have opted to test several wavelets, and then select the most appropriate wavelet on the basis of the performance of the resulting model. Maheswaran and Khosa (2011b) tested different wavelets, order of wavelets and levels of decomposition for various types of time series having diverse features. They selected the best wavelet family for a given time series having specific characteristics. For example, it was found that the Haar wavelet can be recommended for time series having transient features, whereas wavelets such as the db2 and Spline b3 can be

\* Corresponding author. Tel./fax: +1 514 398 7786.

E-mail address: [jan.adamowski@mcgill.ca](mailto:jan.adamowski@mcgill.ca) (J. Adamowski).



recommended for time steps (e.g. monthly) that have long term features. There are other studies where only a single wavelet was used, such as db1, sym8 or db4 (e.g., Kisi (2009)).

Very little research has been published to date that is aimed at establishing general guidelines regarding the choice of wavelet family, in addition to the order of the wavelet for use in wavelet based forecasting models. The general approach that has been adopted to date is to test for different wavelet families and select the best model based on calibration and validation results. However, the intrinsic problem with this approach is that the selected wavelet may work well in capturing certain features of the input time series, but this may not be the case for other features of the input series. For example, a particular wavelet may capture the high flows of a time series very well, but may not perform well in capturing the low flows. Similarly, another wavelet may perform well in capturing the low flow dynamics, but may not be in a position to capture the high flows. This issue has not been investigated in the literature despite its importance, and as such was investigated in this study.

In addition, it is worth noting that the selection of a suitable number of resolution/decomposition levels has also been an issue that has not been resolved, as reported by Maheswaran and Khosa (2011c). In general, the depth or level of decomposition depends on the features of the time series and the nature of the wavelet used. Therefore, it is recommended that instead of using a single best model, one should use an ensemble of forecasts obtained using a group of models developed using different wavelets and different wavelet decompositions. This has also not been explored to date in the literature, and as such was investigated in this study.

Given the necessity of addressing the above issues, this paper proposes a new approach (i.e., the Bayesian Model Averaging based ensemble multi-wavelet Volterra method) that consists of combining ensembles obtained by using different model settings. The underlying idea is that each of these ensembles may account for a particular set of features in the wavelet domain, and may also be able to capture different features (peaks, low flows, time to peak) at the same time. Therefore, by combining the individual forecasts into an ensemble, a richer exploration of the wavelet domain may be achieved, in addition to increasing model performance.

The concept of combining different model setups into an ensemble has been explored in contexts such as daily stream flow forecasting, prediction of rainfall, and hydrological modelling using physically based models (Shamseldin and O Connor, 1999; Georgakakos et al., 2004; Duan et al., 2007; Dhanya and Kumar, 2011; Fundel and Zappa, 2011; Boucher et al., 2011; Zalachori et al., 2012). The basic idea consists of using a set of individual model forecasts instead of the forecasts derived from a single model selected according to a given criterion. These multi model techniques provide a forecast by linearly combining individual model predictions according to different weighting methods. The general approach is to combine ensembles of forecasts that consist of model mean averaging which results in a point forecast. Techniques such as equal weight, Granger-Ramanathan averaging, and Bates-Granger averaging linearly combine the deterministic model outputs into another single point deterministic forecast (Parrish et al., 2012). Shamseldin and O'Connor (1999) explored the use of ANNs for estimating the weights of the different models. However, Hoeting et al. (1999), Raftery et al. (2005), and Rings et al. (2012) determined that these weights do not necessarily reflect the strength of a particular model's performance, and that these techniques are affected by the presence of outliers. Despite this, the multi model ensemble forecasts from these methods were still better than the single best model forecasts. Recently, Bayesian Model Averaging (BMA) has been found to be useful in diverse areas such as statistics, management, meteorology and hydrology.

Bayesian Model Averaging (BMA) is a technique that weighs a model by its performance and likelihood of predicting the observation, resulting in a probabilistic forecast. BMA combines information from a class of models to obtain a probability distribution for a quantity of interest, and accounts for model uncertainty.

Raftery et al. (2005) used the BMA method on a set of meteorological models, while Vrugt and Robinson (2007), Duan et al. (2007) used the BMA technique with hydrologic models. For example, Duan et al. (2007) used BMA to combine different physically based models for hydrologic modelling at a daily scale. The results showed that the BMA based models performed better than the single best model results. More recently, Parrish et al. (2012) used a modified BMA approach that incorporates a sliding window of individual model performances around the forecast. Rings et al. (2012) combined the BMA method with Particle filter and Gaussian mixture modelling with a flexible representation of the conditional probability distribution function. And finally, Zhang and Zhao (2012) improved BMA based Bayesian Neural Networks by combining it with Genetic Algorithms. The above studies demonstrated that the BMA approach is a promising method to combine different models to provide an accurate forecast using an ensemble of forecasts. To date, no research has been published in the literature that explores the use of ensemble coupled wavelet models for short term and long term stream flow forecasting using the Bayesian algorithm. In addition, there are very few papers, if any, that also explore: (i) the selection of the wavelet family; (ii) the order (or length) within the wavelet family; and (iii) the number of wavelet decomposition levels that are appropriate for streamflow forecasting applications. All of these issues are explored in this study for the first time.

The new BMA based ensemble multi-wavelet Volterra method that is proposed in this paper for the first time was tested for stream flow forecasting at monthly, weekly and daily scales. Since the aim of this study was not a comparison of different forecasting methods, but rather to explore, in detail, the above mentioned issues, only the wavelet-Volterra coupled (WVC) method was used as the base model in this study. This particular method was selected since it has already been shown by the first author, in Maheswaran and Khosa (2011c), that the WVC outperforms other new methods that have recently been proposed in the literature for streamflow forecasting (i.e., ANN, wavelet ANN, etc.). The WVC model is simple and versatile and can be implemented in an adaptive mode whereas, in contrast, WANN models are inherently complex and opaque to scrutiny. Further, the proposed WVC based approach yields an analytic form of the forecasting model leading to a better insight into the underlying generating process.

The newly proposed BMA based ensemble multi-wavelet Volterra method is based on the WVC method, but is very different in that it introduces a multi-wavelet BMA based ensemble approach that is coupled with the regular WVC method. In this study, the new BMA based ensemble multi-wavelet Volterra method is compared with the results obtained by individual wavelet-Volterra (WVC) based models that do not have the ensemble algorithm (i.e., the single best WVC model), and the mean averaged ensemble WVC model. The paper is organized as follows. Section 2 provides a brief description of wavelet analysis. The method used for forecasting is described in Section 3. Section 4 presents the BMA methodology. Section 5 describes the study area and the model application. Section 6 provides the validation results of the ensemble models using the BMA technique. Section 7 provides a summary and the conclusions.

## 2. Wavelet analysis

Wavelet analysis is a powerful tool that is used for function analysis similar to Fourier analysis. While Fourier analysis

approximates any periodic function as the sum of sines and cosines by addressing frequency alone, wavelet analysis represents any arbitrary function as the sum of wavelets by addressing both space and scale. The wavelet transform decomposes an arbitrary signal into elementary contributions (or wavelets) which are constructed from one single function called the mother wavelet by adjusting its two parameters, dilation and translation.

For practical applications, the Discrete Wavelet Transform (DWT) is usually preferred. The latter transform is an orthogonal function which can be applied to a finite group of data. Functionally, the DWT has similar attributes to the Discrete Fourier Transform, namely: (i) the transforming function is orthogonal, (ii) a signal passed twice through the transformation is unchanged, (iii) the input signal is assumed to be a set of discrete time samples, and (iv) both transforms are convolutions. However, while the basis function of the Fourier transform is a sinusoid, the wavelet basis, in contrast, is a set of functions that are defined by a localized wavelet function.

A typical discrete wavelet function can be represented as (Labat et al., 2000).

$$\psi_{j,k}(t) = 2^{\frac{j}{2}}\psi(2^j t - k) \tag{1}$$

where  $\psi(t)$  is the mother wavelet and  $j$  and  $k$  are the translation and dilation indices. In the DWT, decimation is carried out so that only half of the coefficients of the detailed component are left at the current level, and half of the coefficients of the smooth version are recursively processed using high pass and low pass filters for coarser resolution levels. Due to the decimation, the number of wavelet coefficients is halved with each move to a coarser level and, as a consequence, there is less information available to train the forecasting model at the coarser level that leads to a reduction in overall forecasting accuracy. This problem, caused by decimation, may be overcome by introducing the stationary or, alternatively, a *trous* wavelet transform proposed by Shensa (1992). The basic idea of the *a trous* wavelet transform is to fill the resulting gaps using redundant information obtained from the original series.

In this approach, the wavelet decomposition is derived by passing the given time series through a low pass filter and, subsequently, from this the derivation of details and the smoothed version becomes possible. For example, consider the original time series  $x(t)$  which may also be denoted by  $c_0$ , or

$$c_0(t) = x(t) \tag{2}$$

Further smoother versions of  $x(t)$  may be derived using (Renaud et al., 2005)

$$c_i(t) = \sum_{l=-\infty}^{\infty} h(l)c_{i-1}(t + 2^{i-1}l) \tag{3}$$

In the preceding Eq. (3),  $i$  takes values from 1 to  $J$  (level of decomposition) and ‘ $h$ ’ is a low pass filter with compact support. The length and characteristics of the low pass filter will depend on the type of wavelet used. The simplest wavelet is the Haar wavelet with a low pass filter specification given by  $(\frac{1}{2}, \frac{1}{2})$ . Similarly, the filter values for the  $B_3$  Spline wavelet are defined as  $(1/16, 1/4, 3/8, 1/4, 1/16)$ . Using the smoother versions of  $x(t)$  at level  $i$  and  $i - 1$ , the detail component of  $x(t)$  at level  $i$  is defined as

$$d_i(t) = c_{i-1}(t) - c_i(t) \tag{4}$$

The set  $\{d_1, d_2, \dots, d_p, c_p\}$  represents the additive wavelet decompositions of data up to the resolution level of  $p$ , and the term  $c_p$  in this set denotes the residual component, which is also called the approximation. Accordingly, for reconstruction, the inverse transform is given by (Renaud et al., 2005)

$$x(t) = c_p(t) + \sum_{i=1}^p d_i(t) \tag{5}$$

Here, unlike the classical DWT, the decimation is avoided, resulting in components at different scales that are of the same length. It should be noted that in this study, the boundary correction was done using the method proposed by Maheswaran and Khosa (2011c).

### 3. Model used for forecasting – Wavelet Volterra Coupled (WVC) Model

As mentioned earlier, all the models used in this study (i.e., single WVC models, mean average ensemble WVC models, and the newly proposed BMA based ensemble WVC models) are based on the wavelet-Volterra coupled model (WVC) developed by Maheswaran and Khosa (2011a,d). The single WVC model was very recently proposed as a new method for hydrological forecasting by the first author of this paper (Maheswaran and Khosa, 2011a, c) for different hydrological forecasting problems such as streamflow forecasting and groundwater level forecasting, and the results showed that the single WVC method is more accurate than other newly developed machine learning methods such as wavelet neural networks, neural networks, and wavelet linear regression. As such, since the WVC method has already been shown to be more accurate than other new methods for hydrological forecasting, this paper focused on improving the WVC method, in addition to exploring several additional important issues that have not been addressed in the literature regarding the use of wavelet transform based models in hydrological forecasting applications. Since the aim of this study was not to compare different forecasting methods, a common reference model was considered (the single WVC model), and the performance of the newly proposed Bayesian Model Averaging based ensemble multi-wavelet Volterra method was compared with the base model (i.e., single WVC model), as well as the mean average ensemble WVC model.

Fig. 1 presents a general schematic of the structure of the WVC forecast model design. As depicted in the figure, an input signal,  $Y = (y_1, \dots, y_{n-1})$ , is decomposed to obtain wavelet coefficients at different scales using the *a trous* wavelet transform, and the resulting wavelet decomposition at various levels are then appropriately integrated using the second order Multiple Input Single Output (MISO) Volterra formulation. The Volterra series representation of a nonlinear time-invariant system with memory is based on a simple extension of the Taylor series expansion for nonlinear autonomous causal systems with memory. Studies such as Diskin and Boneh (1972), Amorocho (1973), and Labat and Ababou (2001), have led to the general understanding that a second order representation (or, a second order nonlinear Volterra kernel) is usually sufficient for most hydrologic systems. For additional mathematical details, reference may be made to the accompanying appendix.

To understand the formulation, let  $u_1, u_2, \dots, u_J$  denote the wavelet coefficients at each scale, and let the scaling coefficients be denoted as  $u_{J+1}$ , where  $J$  is the coarsest level of decomposition. The wavelet coefficients and scaling coefficients of the original series are nonlinearly convolved using the second order Volterra representation within a multiple inputs-single output framework. If  $J$  denotes the level of decomposition,  $N$  is the number of inputs,  $m$  denotes the memory length at each level,  $\xi_t$  represents the model noise (including modelling errors and the unobservable disturbances), and the multiscale nonlinear relationship may be written as

$$y(t) = \sum_{n=1}^{J+1} \sum_{\tau=1}^m h_1^{(n)}(\tau) u_n(t - \tau) + \sum_{n=1}^{J+1} \sum_{\tau_1=1}^m \sum_{\tau_2=1}^m h_{2s}^{(n)}(\tau_1, \tau_2) u_n(t - \tau_1) u_n(t - \tau_2) + \sum_{n_1=1}^{J+1} \sum_{n_2=1}^{n_1-1} \sum_{\tau_1=1}^m \sum_{\tau_2=1}^m h_{2x}^{(n_1, n_2)}(\tau_1, \tau_2) u_{n_1}(t - \tau_1) u_{n_2}(t - \tau_2) + \xi_t \tag{6}$$

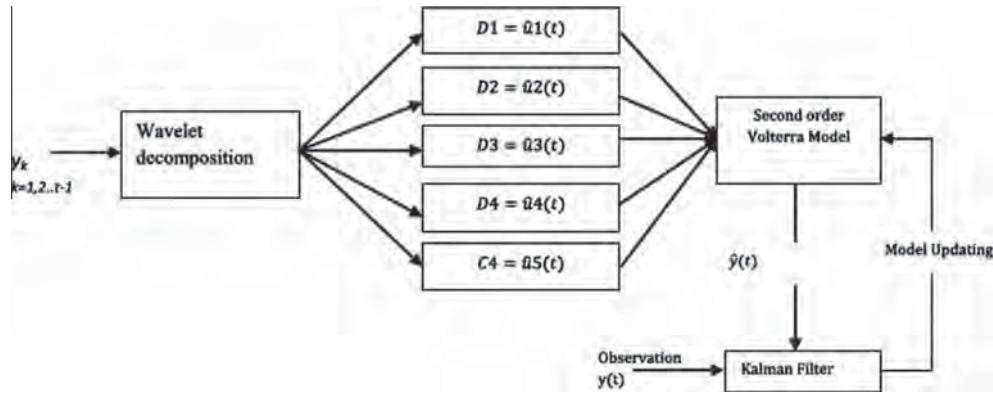


Fig. 1. Wavelet Volterra Coupled (WVC) model.

In Eq. (6), the first order kernels  $h_1^{(n)}$  describe the linear relationship between the  $n$ th input  $u_n$  and the output signal  $y$ . The second order self-kernels  $h_{2s}^{(n)}$  describe the 2nd order nonlinear relation between the  $n$ th input  $u_n$  and  $y$ , respectively, and the second order cross-kernels  $h_{2x}^{(n_1, n_2)}$  describe the 2nd order nonlinear interactions between each unique pair of inputs ( $u_{n_1}$  and  $u_{n_2}$ ) as they affect  $y$ . Eq. (6) can be simplified by combining the last two terms to yield Eq. (7) (it now remains to estimate kernels  $h_1$  and  $h_2$ )

$$y(t) = \sum_{n=1}^{J+1} \sum_{\tau=1}^m h_1^{(n)}(\tau) u_n(t - \tau) + \sum_{n_1=1}^{J+1} \sum_{n_2=1}^{J+1} \sum_{\tau_1=1}^m \sum_{\tau_2=1}^m h_2^{(n_1, n_2)}(\tau_1, \tau_2) u_{n_1}(t - \tau_1) u_{n_2}(t - \tau_2) + \xi_t \quad (7)$$

The representation of Eq. (7) can be further simplified by considering each of the lagged variables  $u_1(t - 1), u_1(t - \tau), \dots, u_2(t - 1), u_2(t - \tau), \dots$  as separate variables  $d_1(t), d_2(t), d_3(t), \dots, d_{N_l}(t)$ . Eq. (7) can then be written as (Maheswaran and Khosa 2011c)

$$y(t) = \sum_{l=1}^{N_l} h_1(l) d_l(t) + \sum_{l_1=1}^{N_l} \sum_{l_2=1}^{N_l} h_2(l_1, l_2) d_{l_1}(t) d_{l_2}(t) \quad (8)$$

More clearly

$$d_l(t) = \{x_k(t) \mid 1 \leq k \leq J + 1; 1 \leq l \leq J + 1\}$$

$$d_l(t) = \{x_k(t - \tau), 1 \leq k \leq J + 1; J + 1 < l \leq N_l; \tau = 1, 2, 3 \dots m\}$$

$\tau = \tau$ th lagged value.

$J =$  level of decomposition.

$N_l =$  total number of lagged variables

In the preceding Eq. (8),  $h_1$  and  $h_2$  represent the Volterra kernels to be estimated, and it can be seen that the number of parameters to be estimated increases proportionately as the number of inputs and/or the process memory increases, leading to an increased computational burden and severely compromised estimation. In order to handle these computational difficulties, Chen et al. (1989) and, later, Wei and Billings (2004) proposed the use of the Orthogonal least squares technique as a preferred estimation approach as it is better, in comparison with the Ordinary Least Squares technique, at handling modelling situations where there is a possibility of multi-collinearity amongst vectors that constitute the coefficient matrix.

In addition to the above, as shown in Fig. 1, the proposed formulation is recursively updated in real time using the well known Kalman Filter formulation. For further derivation of the single

WVC model formulation (which forms the basis of both the mean average WVC ensemble method, as well as the new BMA based ensemble WVC method, described in subsequent sections), readers are referred to Maheswaran and Khosa (2011c).

#### 4. Ensemble combination

There are different models that researchers can use in data based hydrological forecasting and the choice between them is often not obvious. Thus, the uncertainty over which model to use is an important issue in forecasting. Since model uncertainty critically affects forecasts (and forecast uncertainty), the ‘pooling’ or combination of forecasts using different models is one approach to address this issue. This pooling can be done in a variety of different ways. The simplest method is the mean averaging of the ensembles (in our study this is the average ensemble WVC model), wherein the weights in combining different forecasts is uniform. Further, advanced techniques can be based on historical performance (Hendry and Clements, 2002). A related methodology for dealing with large numbers of regressors used in macroeconomic forecasting is based on principal components or factors (e.g., Stock and Watson, 2002). Alternatively, the weights used in combining forecasts can be based on posterior model probabilities within a Bayesian framework. This procedure, which is typically referred to as Bayesian Model Averaging (BMA), is the standard approach to model uncertainty within the Bayesian paradigm, where it is natural to reflect uncertainty through probability (Steel, 2011). Thus, it follows from the direct application of Baye’s theorem, as is explained in Leamer (1978), Min and Zellner (1993), Raftery et al. (1997), Fernandez et al. (2001) and Ley and Steel (2009), that pooling or combining models using BMA minimizes the expected predictive squared error loss, provided that the set of models under consideration is exhaustive.

This study explored the use of two ensemble methods, namely: the mean average ensemble WVC method, and the newly proposed BMA based ensemble WVC method, and compared these two approaches to the single WVC method.

##### 4.1. Bayesian Model Averaging

Statistical analysis usually assumes that there is a “single best” model, often selected from a class of several possible models. It is typically not the case that the selected model is always the best, and the analysis ignores model uncertainty and assumes that the data obtained from the selected model has no uncertainty. This approach results in overconfident inference and forecasts, especially when substantially different scientific results may be obtained from alternative models (see Hoeting et al., 1999). Bayesian Model

Averaging reduces the potential overconfidence by conditioning not on a single model, but on a class of models. Bayesian Model Averaging (BMA) is a statistical approach for post-processing ensemble forecasts from multiple competing models (Leamer, 1978). It has been widely used in a variety of areas, as mentioned earlier, such as dynamical weather forecasting models (Raftery et al., 2005) and hydrology. For example, in hydrology, Ajami et al. (2007) and Duan et al. (2007) used the BMA approach to provide an ensemble of three different hydrologic models. In their work, they used BMA to combine the model results from three different physically based lumped hydrologic models. They showed that combining the model results via BMA provided better results than the individual models.

The basic principle of the BMA method is to generate an overall forecast probability distribution function (PDF) by taking a weighted average of the individual model forecast PDFs. The weights represent the model performance, or more specifically, the probability that a model will produce the correct forecast. The BMA method assumes that the probability of an observation  $y_{obs}(t)$  at time  $t$  is given by a weighted sum over a number of probability distributions  $g(y_f(k,t))$  from the  $k$  different forecasting models

$$P(y_{obs}(t)) = \sum_k w(k)g(y_f(k,t)) \quad (8)$$

The forecast distribution of each individual model is assumed to be a normal distribution with variance  $\sigma(k)$

$$g(y_f(k,t)|y_{obs}(t), \sigma(k)) = \frac{1}{\sigma(k)\sqrt{2\pi}} \times \exp\left(-\frac{(y_{obs}(t) - y_f(k,t))^2}{2\sigma(k)^2}\right) \quad (9)$$

The BMA algorithm finds the optimal values for  $w(k)$  and  $\sigma(k)$ , such that the likelihood of the overall PDF (Eq. (9)) is maximal, given a set of historical forecasts and observations. The method does so by optimizing  $w(k)$  and  $\sigma(k)$  consecutively in an iterative scheme. The first step of the iteration starts with an initial guess for the weights  $w(k)$  and  $\sigma(k)$  for each of the individual models and estimates the matrix  $z(k,t)$  using Eq. (10), which represents the probability that model  $k$  gives the best forecast for station  $s$  at time  $t$

$$z(k,t) = \frac{g(y_{obs}(t)|y_f(k,t), \sigma(k))}{\sum_k g(y_{obs}(t)|y_f(k,t), \sigma(k))} \quad (10)$$

The second step in the iterative algorithm is to determine the weights  $w(k)$  and variances  $\sigma(k)$  of each of the model's  $k$ , based on the values of  $z(k,s,t)$ . The weights are estimated using

$$w(k) = \frac{1}{n} \sum_t z(k,t) \quad (11)$$

where  $n$  is the number of observations in the training period ( $t$ ). The variance  $\sigma(k)$  is estimated using

$$\sigma^2(k) = \frac{\sum_t z(k,t)(y_{obs}(t) - y_f(k,t))^2}{\sum_t z(k,t)} \quad (12)$$

The two steps are alternated to convergence; that is, when  $w(k)$  and  $\sigma(k)$  no longer change after a recalculation of  $z(k)$ . One can use a convergence criterion or alternatively use a fixed number of iteration cycles that should guarantee convergence. In this study, the Expectation-Maximization (EM) algorithm (Raftery et al., 2005) was used for the purpose of optimization. For further details on the EM algorithm, readers are referred to Raftery et al. (2005).

For dynamical model applications, one can use the weights and variances from the previous time step as a starting point for the new iteration.

#### 4.2. Performance evaluation of ensemble forecasts

The performance of forecasting models can be evaluated in terms of goodness of fit after each of the model structures is calibrated. The coefficient of determination ( $R^2$ ) measures the degree of correlation among the observed and predicted values.  $R^2$  values range from 0 to 1, with 1 indicating a perfect relationship between the data and the line drawn through them, and 0 representing no statistical correlation between the data and a line. The Root Mean Square Error (RMSE) evaluates the variance of errors independently of the sample size. RMSE indicates the discrepancy between the observed and forecasted values. A perfect fit between observed and forecasted values would have a RMSE of 0. The Nash–Sutcliffe model efficiency coefficient (NSC) can also be used to assess the forecasting power of hydrological models. An efficiency of one (NSC = 1) corresponds to a perfect match of forecasted data to the observed data. An efficiency of zero (NSC = 0) indicates that the model predictions are as accurate as the mean of the observed data.

In order to evaluate the quality of ensemble forecasts, the rank histogram (or Talagrand diagram (Talagrand et al., 1999)), can be used. It is a useful tool that allows one to assess the calibration of the predictive distribution. To construct it, the observed time series  $x_t$  value is added to the ensemble forecast. That is, if the forecast has  $n$  members the new set consists of  $n + 1$  values. Then, the rank associated with the observed value is determined. This operation is repeated for all  $N$  forecasts and corresponding observations in the archive. The rank histogram is obtained by constructing the resulting  $N$  ranks. If the predictive distribution is well calibrated, then the rank histogram should be close to flat (Boucher et al., 2009). An asymmetrical histogram is usually an indication of a bias in the mean of the forecasts. If the rank histogram is symmetric and U shaped, it may indicate that the predictive distribution is under dispersed. If it has an arch form the predictive distribution may be over dispersed.

## 5. Model application

### 5.1. Study area

The river flow data used in this study were obtained from the US Geological Survey (USGS). The time series of daily stage data from two stations were used: (i) 12414500 Selway River Nr Lowell ID, USA; and (ii) 13336500 ST Joe River at Calder ID, USA. The above two stations were selected because the flow regimes in both rivers are highly varied, and are therefore useful in testing the capability of the models in a rigorous manner (low flows, high flows). The catchment areas and also the flow statistics are provided in Table 1.

Further information on these two stations can be found on the USGS web server (<http://www.usgs.gov>). Since the forecasting in this study had three different lead times (daily, weekly and monthly), the observed data (half hourly sampled data) for a period of 45 years was used and the daily, weekly and the monthly mean for these stations were accordingly estimated. 60% of the entire data sets were used for calibration and the remaining data was used for validation purposes.

### 5.2. Model development – proposed BMA based ensemble multi-wavelet Volterra model

#### 5.2.1. Generation of ensembles

For the purpose of creating an ensemble of forecasts, several WVC models were run by varying the inputs. This was done by



**Table 1**  
Statistics of streamflow data observed at the two stations.

Station no	USGS ID	State	Area (sq km)	Ele (m)	Mean (m <sup>3</sup> /s)	SD (m <sup>3</sup> /s)	Max. (m <sup>3</sup> /s)	Min. (m <sup>3</sup> /s)
I	12414500	ID	2668	1283	66.8	76	393.6	6.7
II	13336500	ID	4947	1719	106.2	135.6	690.1	9.2

varying the mother wavelets, length (order) of wavelet within that wavelet family, number of decomposition levels, and the choice of the combination of input wavelet coefficients. The wavelet transform of the flow series were calculated with different filters of the Daubechies (db1,db2, db4,db5, db6,db7), Symlet (sym2, sym4) and Spline (b3-Spline) wavelet families, which are commonly used in wavelet based forecasting (e.g., Nourani et al., 2009; Kisi 2009 and Maheswaran and Khosa 2011c). Overall, 10 filters were employed. In each case, the number of decomposition levels was varied from one to a maximum of  $J$ . It is worth noting that the maximum number of decomposition levels depends on the filter length and the time series length. However, the upper limit for the number of decomposition levels was set as 6 for monthly time series, 8 for weekly time series and 10 for daily time series based on previous research (Maheswaran and Khosa, 2011c), where it was observed that increasing the level of decomposition beyond a certain limit deteriorates model performance. The initial set of input wavelet coefficients to be used as the inputs for the models were determined using correlation analysis. The correlation analysis shows the influence of the different decompositions on the original signal. Only wavelet coefficients with correlation coefficients above 0.6 were selected for modelling. The selected inputs were then used to form different combinations of the wavelet coefficients from the different levels of decomposition (note: only within the decomposition from the same wavelet).

The above approach results in a number of input combinations which are then used as inputs for the WVC models used to produce the ensemble of forecasts. In this study, for all the WVC model based approaches (i.e. single, mean average, and BMA WVC models), the stream flow time series is decomposed into wavelet ( $D_i, D_{i+1} \dots D_N$ ) and scaling ( $A_N$ ) coefficients at different scales. The approximation and the detail components represent the slow and fast components that make up the integrated history of the time series of observed flows. These components are combined through the Volterra framework to forecast future flows followed by a Kalman Filter based updating procedure that uses the newest (current) observation to update the model for use at the next time step for further forecasts. Once the WVC models are calibrated for the different input variable combinations, these models are used to produce an ensemble of forecast values.

#### 5.2.2. Single best WVC model

From the set of numerous model runs, the model having the best performance in terms of NSC and Root Mean Square Error (RMSE) was selected and used for forecasting during the validation period.

#### 5.2.3. Ensemble combination by mean averaging

The ensembles of forecasts obtained from the different WVC model setups were combined using mean averaging, and a single point forecast was produced. This was the mean average WVC ensemble model.

#### 5.2.4. BMA WVC approach

From these ensembles of forecasts, only those forecasts that had a certain Nash Sutcliffe Criteria (NSC above 0.75) were selected to form the final set of ensembles. Then the forecast results from these selected models were combined using the BMA technique.

The BMA algorithm assigns suitable weights to each of the model setups based on the a priori performance of these models in the calibration or training stage. These weights are then assigned for the corresponding model forecasts during the validation stage to obtain the BMA based forecasts.

Fig. 2 shows the entire schematic of the proposed BMA based ensemble multi-wavelet Volterra method. This approach was adopted for developing models at the different temporal scales (monthly, weekly and daily).

## 6. Results and discussion

In this section, the results of the different models for forecasting stream flow at monthly, weekly and daily time scales using the single WVC models, the average WVC ensemble models, and the BMA WVC ensemble models are reported.

### 6.1. Model results at a monthly scale

Varying the wavelets, order of wavelets, decomposition levels, and the input variables resulted in numerous WVC models which were in turn used for producing the ensemble of forecasts. Table 2 provides a detailed report on the sensitivity analysis of the different wavelets, decompositions and input combinations across different scales for the single WVC model.

The analysis of the results from these different WVC model setups provided some interesting findings.

1. It seems that not all of the wavelet filters were able to provide good forecasts. For example, it was observed that higher order wavelets such as db5 and sym4 decreased model performance when compared to models based on lower order wavelets such as Haar, db2 and sym2.
2. It was found that increasing the depth of decomposition (i.e., number of decomposition levels) in some cases (using the 'Daubechies' wavelet and 'Haar' wavelets) increased performance but only to a certain level. For example, in the case of the Haar wavelet the model performance using 4 levels of decomposition was better than that of the model using 3 and 5 levels of decomposition. However, in cases with 'Symlet' and higher order 'Daubechies' wavelets, the performance of the model decreased significantly when the level of decomposition was increased.
3. It was also observed that the different wavelets captured different features of the time series. Fig. 3 shows the hydrograph of the results from the Haar wavelet based WVC model. It can be seen that the model captured the low flows very well, whereas it could not capture the peak flow. However, the model results from the Symlet and Spline wavelet based WVC model (Figs. 4 and 5, respectively) show that the Symlet wavelet captures the peaks (over estimation), whereas it does not perform well for low flows.

The above analysis shows that the performance of the single WVC model depends on the choice of the wavelet, and also on the level of decomposition. It can also be seen that a 'one best choice' of the model set up is not ideal, and instead a combination of the forecasts from different models would improve the overall

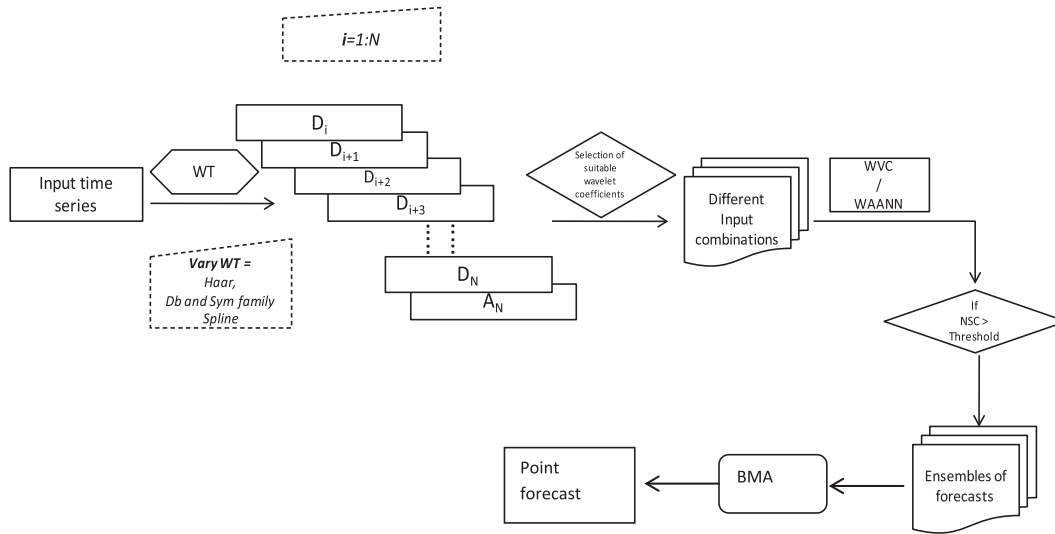


Fig. 2. Proposed BMA WVC algorithm.

Table 2  
Summary of model results from various single WVC models for one step ahead forecasting at a monthly scale.

Wavelet used	Level of decomposition	Memory (m) at each level {DW <sub>i</sub> 's, C}	RMSE (m <sup>3</sup> /s)	MAE (m <sup>3</sup> /s)	NSC
<i>Station I</i>					
B3-Spline	4	{2,2,2,2,1}	58.82	33.42	0.57
Db4	4	{2,2,1,1,2}	56.6882	32.43	0.73
Db5	3	{1,1,1,2}	56.53	32.85	0.74
Haar	3	{2,2,2,1}	51.30	30.24	0.78
Db2	4	{1,2,2,2,1}	54.93	32.78	0.75
Sym4	5	{1,1,1,2,2,2}	62.27	30.47	0.64
Sym4	6	{1,2,1,1,2,2,2}	84.99	46.5	0.54
Haar	4	{2,2,2,2,1}	57.35	33.43	0.74
<i>Station II</i>					
B3-Spline	3	{2,2,2,1}	37.88	19.88	0.85
Sym4	4	{3,2,2,2,1}	37.01	19.00	0.88
Db2	3	{1,1,1, 2, 1}	36.88	18.78	0.89
Haar	3	{1,2,2,2}	37.85	19.7	0.87
Haar	4	{1,1,2,2,2}	36.54	18.35	0.90
Haar	5	{1,1,1,2,2,2}	40.45	21.78	0.54

performance of the model. Therefore, in order to improve the forecasts, instead of taking one best model, the ensembles from different models were combined using: (i) the mean averaging approach and (ii) the newly proposed methodology based on BMA. The BMA technique provides the weights to be assigned to the forecasts from each model setup according to the overall performance of the model during the training stage. For both the stations, a total of 140 model runs were performed, and the best 45 models were

selected after applying the threshold of 0.75 (for NSC). The rank histogram for these ensembles is shown in Fig. 6. It can be seen that the histogram for Station I is neither flat nor U shaped, indicating that there is a possible under dispersion of the predictive distributions. However, for Station II, the histogram is relatively flat, indicating that the predictive distribution is properly calibrated.

The weights of the corresponding models obtained via the BMA technique for the two stations are shown in Fig. 7a and b. The

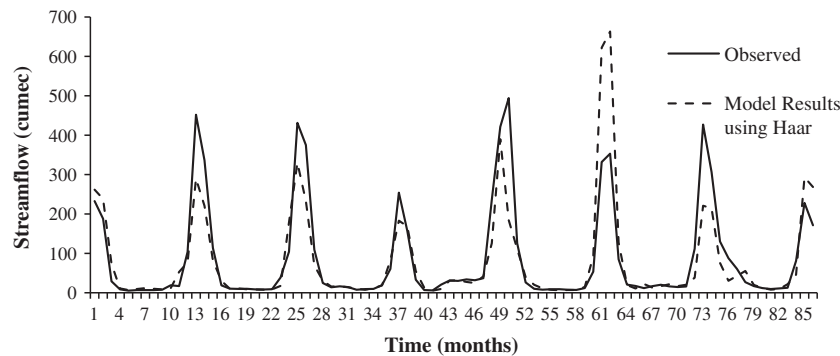


Fig. 3. Model results for single WVC model using Haar wavelet at Station I.

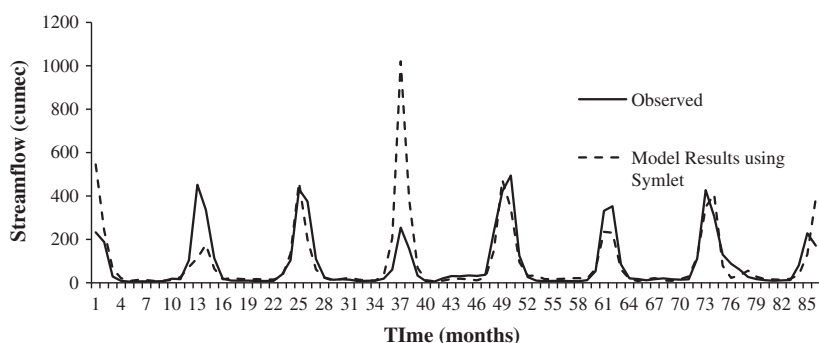


Fig. 4. Model results for single WVC model using Symlet wavelet at Station I.

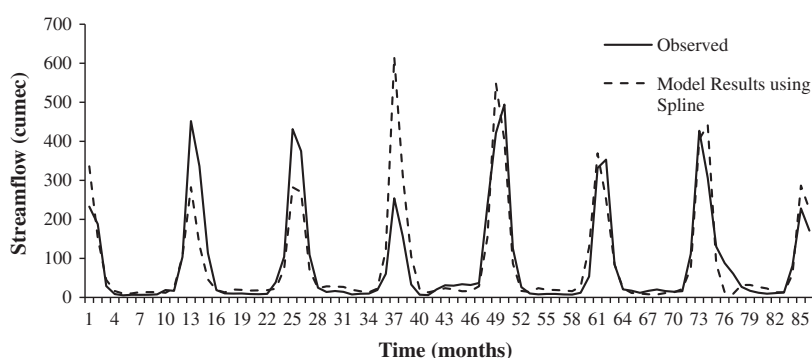


Fig. 5. Model results for single WVC model using Spline wavelet at Station I.

corresponding wavelets and the decomposition levels used in the BMA based ensemble multi-WVC model are also shown in Fig. 7a and b. It can be seen that the Haar wavelet, db2, db4 wavelets and sym4 wavelets contribute significantly to the forecast accuracy. The model results are also plotted in the form of a hydrograph in Figs. 8a and b. It can be seen that the ensemble model based on BMA outperforms the single best WVC model, as well as the average ensemble WVC model. The BMA based ensemble model is able to capture the low and high flows more closely than the other models. The results for Station 2 are also shown in Fig. 8b. Table 3 compares the one month ahead forecast statistics from the best single model, the average ensemble WVC model and the BMA based ensemble WVC model. The results show that using the BMA ensemble method results in higher model accuracy and also higher reliability in the forecast compared to that of the single best model and the mean average ensemble WVC model.

### 6.2. Model results at a weekly scale

Table 4 shows the results of a few of the single WVC models. Fig. 9a shows the hydrograph of the results from the Haar wavelet (at level 3 decomposition) based single WVC model. It can be seen that the model captured the low flows very well, whereas it could not capture the peak flow. However, the model results from the Symlet and Spline single WVC model (Fig. 9b and c, respectively) show that the Symlet wavelet captures the peaks (over estimation), whereas it does not perform well for low flows. It can be seen that the single WVC models based on Symlet wavelets do not perform well in comparison with the Haar, db2 and db4 based models.

Also, it is to be noted that if the decomposition is performed to a level of 3, the single WVC model is able to pick up the low flows better (see Fig. 9a and d). However, when a 4th or 5th level

decomposition is applied, the model is able to capture the underlying seasonality and high flows whereas the low flows are over estimated (see Fig. 9b and c). From these set of models (the best single WVC models), the average WVC ensemble models were formed.

For the BMA WVC models, from the total set of 75 models which were evaluated, those having a NSC  $\geq 0.75$ , (i.e. only 39) were selected for ensemble formation. The rank histogram of these ensembles is shown in Fig. 10. It can be seen that for Station I, the histogram is more or less flat indicating the good calibration of the predictive distribution (Boucher et al., 2009), whereas for Station II, the rank histogram is U-shaped, indicating the under dispersive nature of the predictive distribution. However, Hamil (2001) argues that the U-shaped behaviour may be due to selecting ensemble members with equal prediction probabilities.

The corresponding weights of the BMA WVC models are shown in Fig. 11a (for Station I). Similarly, for Station II the same approach was repeated and the weights of the models were obtained from the BMA algorithm as shown in Fig. 11. The results of the BMA based WVC model, along with the best single WVC model, and the mean average ensemble WVC model, are shown in Table 5. For both the stations, it can be seen that the BMA based WVC model performs better than the best single WVC model, as well as the mean average ensemble WVC model. Figs. 12a and b show the model results of the validation data set for these three models, and it can be observed that the BMA based WVC ensemble model performs better in capturing the peaks as well as the low flows.

### 6.3. Model results at a daily scale

At the daily scale, a similar approach was used as in the monthly and weekly time scales to develop the single WVC models, the mean average WVC ensemble models, and the BMA WVC

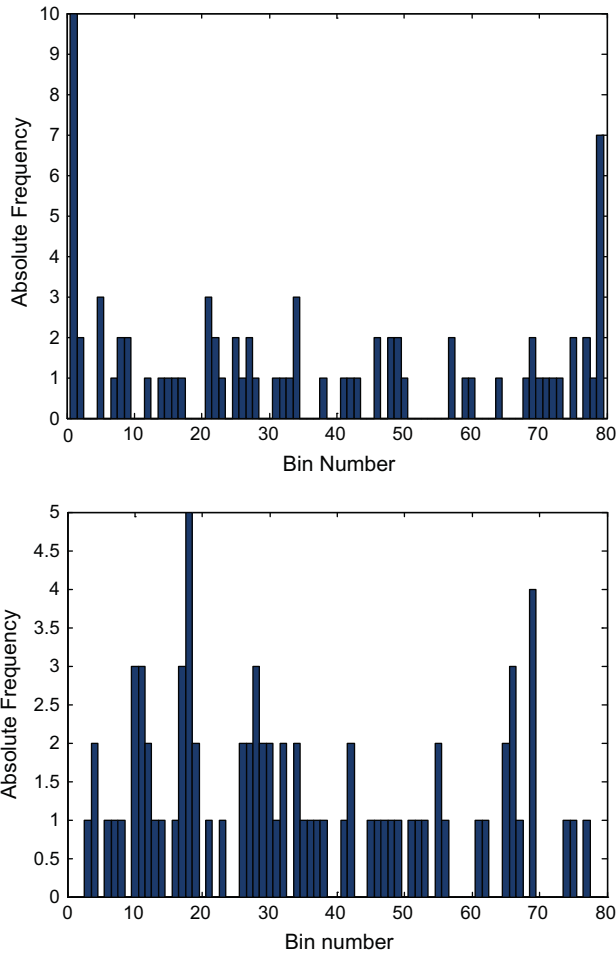


Fig. 6. Rank Histogram of the ensembles obtained from WVC models for (a) Station I and (b) Station II.

ensemble models. Table 6 shows the results of the various single WVC models.

Analysis of these single WVC model results and the corresponding model set ups revealed the following information:

1. It was observed that the lower order wavelet filters such as Haar and db2 provided better forecasts than those obtained using the higher order wavelet filters in the case of daily time scale forecasting. This observation is in congruence with the results of Nourani et al. (2009), who showed that lower order filters performed better than higher order filters and longer wavelets. In our study, it was observed that the db4 and Symlet wavelets performed poorly in comparison with the Haar wavelet. This may be attributed to the fact that the db4, sym4 and sym8 wavelets have broader length (higher order), and are therefore not able to capture the peak flows accurately.
2. With regards to the depth of decomposition (i.e., decomposition level), it was observed that increasing the depth drastically reduced the model performance. It was determined that the maximum depth of decomposition is 2.

A total of 40 different single WVC models were run and 22 models having a threshold of NSC  $\geq 0.90$  were selected for combining the forecasts using the BMA WVC model. Figs. 12a and b show the weights that were applied to the model. It can be seen that the Haar wavelet model's contribution is very significant in comparison with the other models. The model results for Station I are in the form of a scatter plot for better clarity (Fig. 13). It can be seen that the BMA algorithm significantly improved WVC model performance when compared to the single WVC model and the mean average WVC ensemble model. Similar analysis was done for Station II and the summary of results obtained is provided for both the stations in Table 7.

It can be seen from the summary of the results that there was a significant improvement in the model performance when the BMA WVC model was used in the case of Station I, compared to using the best single WVC model. However, in the case of Station II, the improvement in model performance was not very significant. This

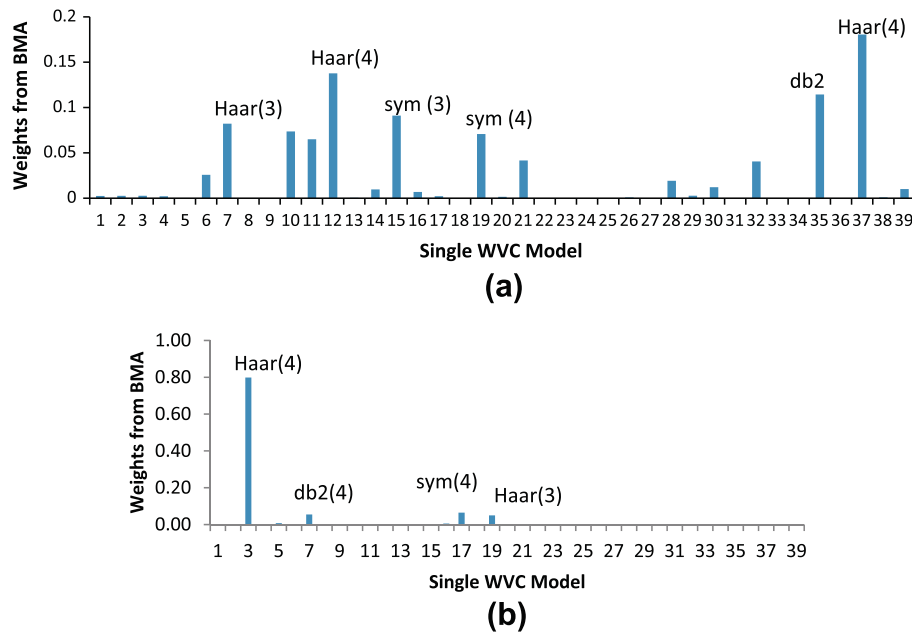


Fig. 7. A single set of BMA weights computed over the entire training period.



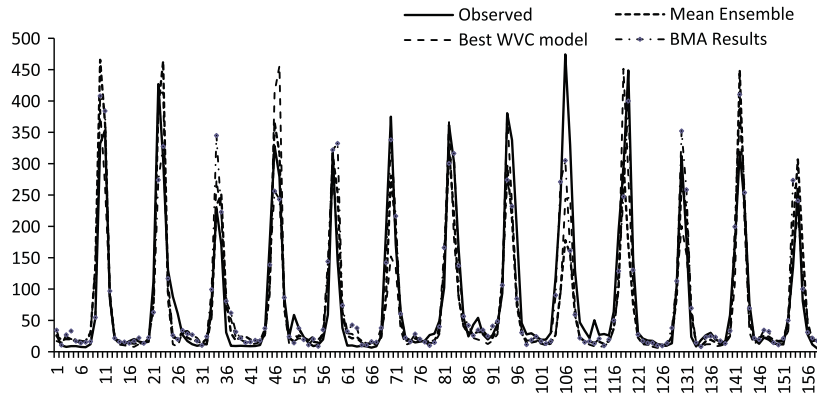


Fig. 8a. Validation results for Station I at a monthly scale.

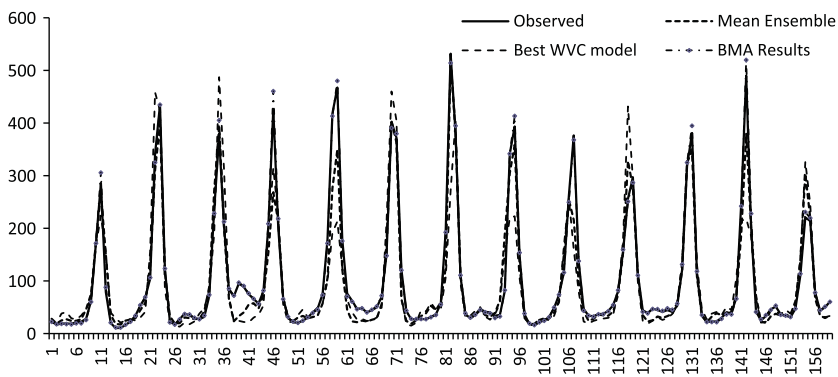


Fig. 8b. Validation results for Station II at a monthly scale.

**Table 3**  
Results for the different models for Stations I and II.

Performance measures	Best single wavelet Volterra (WVC)	Average WVC ensemble	BMA based WVC ensemble
<i>Station I</i>			
RMSE (m <sup>3</sup> /s)	51.30	53.69	42.54
MAE (m <sup>3</sup> /s)	30.24	28.92	24.56
Correlation coefficient	0.89	0.88	0.92
NSC	0.78	0.77	0.85
<i>Station II</i>			
RMSE (m <sup>3</sup> /s)	36.54	34.56	26.56
MAE (m <sup>3</sup> /s)	18.35	19.26	13.56
Correlation coefficient	0.94	0.96	0.98
NSC	0.90	0.91	0.95

may be due to the fact that all the models in the ensemble formation may be having the same prediction probabilities. Overall, it can be seen that using the BMA technique increased WVC model performance and reduced the uncertainty in the predictions (see Fig. 14).

**7. Summary and conclusions**

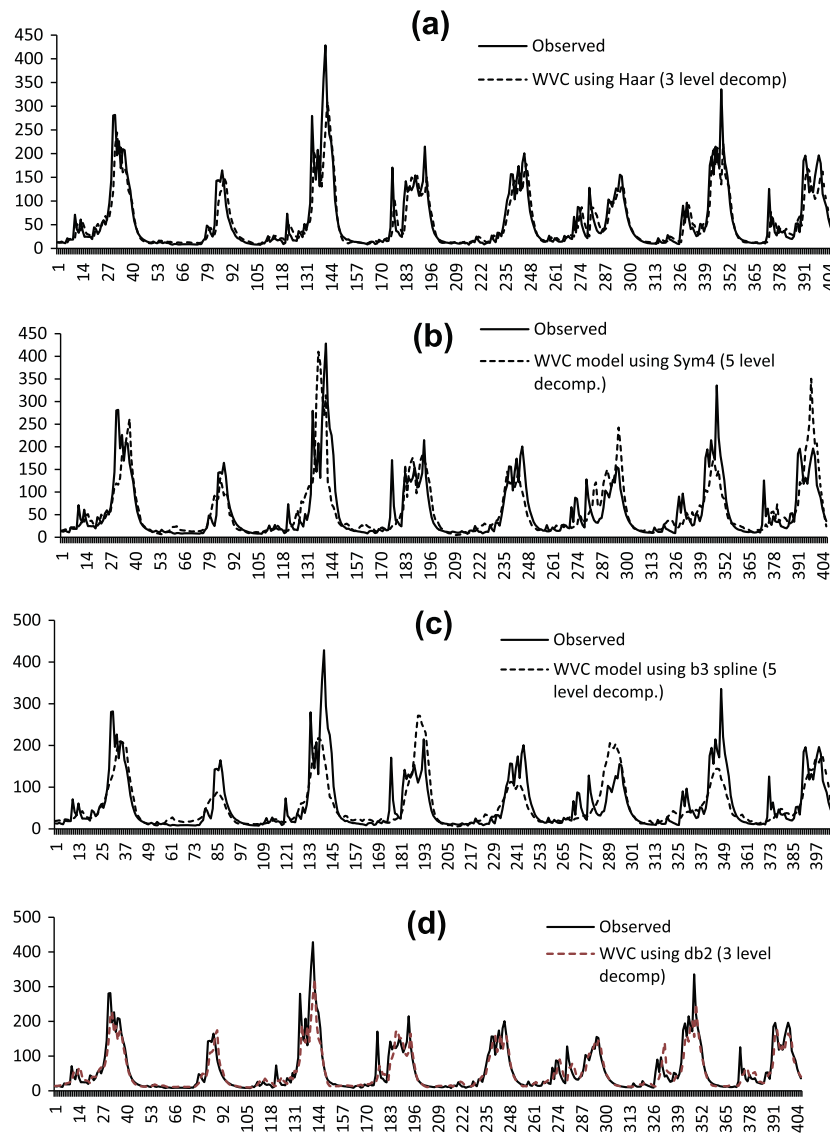
Different wavelets have unique properties and thus possess different strengths to capture specific aspects of streamflow time series. Therefore, it is highly desirable to have a combined forecasting model that possesses the advantages of different wavelets. In this study, a new wavelet ensemble forecasting approach (i.e., the BMA based ensemble multi-wavelet Volterra approach) that takes advantage of different properties of multiple wavelets was proposed and tested for the first time. This was accomplished through

a weighting strategy based on Bayesian Model Averaging. The major findings of this study are:

1. Different wavelets having unique properties possess the ability to capture specific features of real world processes such as streamflows. In this study, it was seen that in the monthly time series, the Haar wavelet and the db wavelet performed well in capturing the low flows, whereas the Symlet wavelet captured the high flows and time to peaks more accurately. It was observed that at a weekly time scale, the Haar, db2, and db4 wavelet based models produced better results than the Symlet wavelet based models. However, in the case of the daily scale, the Haar and db2 wavelet based models outperformed the other wavelet based models for both high and low flows. This issue has not really been explored to date in the literature in any

**Table 4**  
Different single wavelet Volterra models for one step ahead forecasting at a weekly scale.

Wavelet used	Level of decomposition	Memory (m) at each level {DW <sub>i</sub> 's, C}	RMSE (m <sup>3</sup> /s)	MAE (m <sup>3</sup> /s)	NSC
<i>Station II</i>					
B3-Spline	3	{2,2,2,1}	55.08	16.89	0.75
B3-Spline	4	{1,2,2,2,1}	58.82	29.91	0.68
B3-Spline	5	{2,2,2,2,2,1}	65.85	32.54	0.58
db2	4	{2,2,1,1,2}	56.4	29.45	0.72
db2	3	{1,1,1,2}	53.15	26.54	0.76
db2	4	{1,2,2,2,1}	56.45	29.04	0.70
Haar	3	{1,1,1,2}	54.56	26.95	0.75
Haar	3	{2,2,2,1}	52.14	26.43	0.79
Sym4	4	{2,2,2,2,2}	57.45	27.45	0.74
Sym4	5	{1,1,1,2,2,2}	58.45	29.82	0.69
<i>Station I</i>					
B3-Spline	3	{2,2,2,1}	35.08	18.97	0.77
Db5	4	{3,2,2,2,1}	39.02	20.56	0.65
Sym4	3	{1,1,2,1}	34.65	16.7	0.78
Haar	3	{1,2,2,2}	33.78	16.07	0.79
Db4	4	{1,1,2,3,3}	38.2	18.69	0.68
Db2	3	{1,2,2,2}	34.35	16.73	0.79



**Fig. 9.** Model results for WVC model using (a) Haar wavelet, (b) Sym 4, (c) Spline and (d) db2 wavelet at Station I.

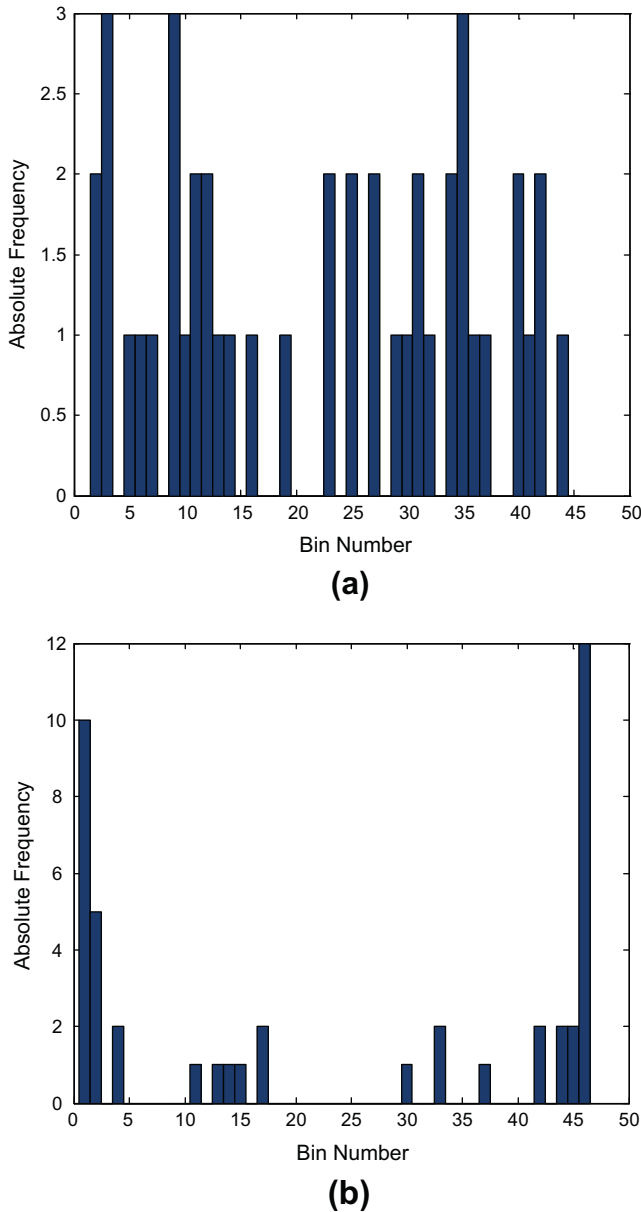


Fig. 10. Rank Histogram of the BMA–WVC ensembles for (a) Station I and (b) Station II.

detail, so the results of this study should prove to be very useful in helping to select appropriate wavelets for different time scales in streamflow forecasting.

2. It was found that simply averaging the model results by assigning equal weights to the model (i.e. the mean average ensemble WVC approach) does not improve the results in comparison with the best single model. In some cases, averaging decreases the performance of the model. However, weighted averaging based on the BMA technique improved model performance significantly. In the case of monthly forecasting, the BMA based WVC model improved the forecasting performance by 9% and 20% with respect to the best single WVC model for Stations I and II, respectively. Similarly, in the case of the weekly forecasts, the BMA based WVC model improved the forecasting performance by 3.85% and 14% for Stations I and II, respectively. In the case of daily forecasts, the BMA based WVC model produced

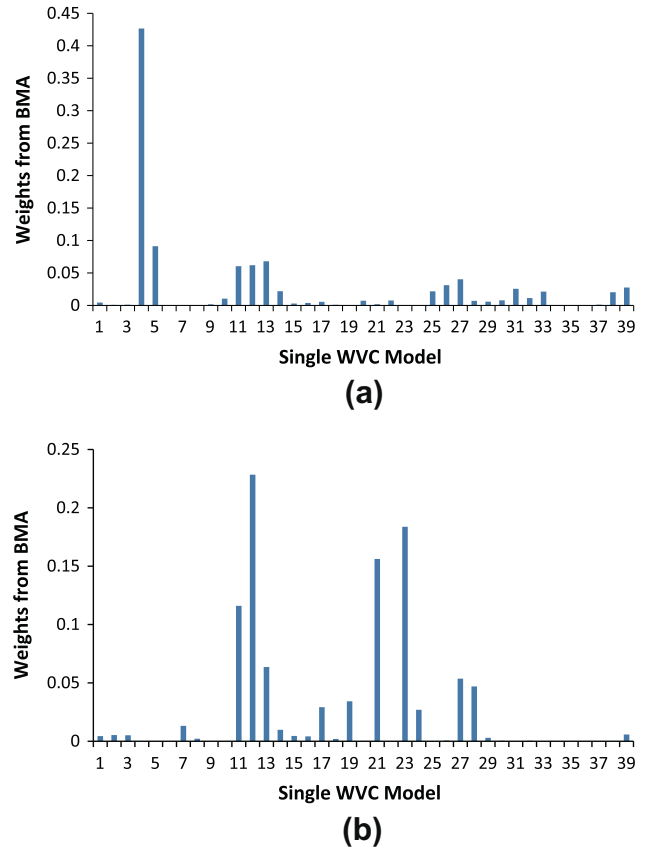


Fig. 11. A single set of BMA weights computed over the entire training period for the BMA WVC models at a weekly scale at (a) Station I and (b) Station II.

slightly better results than the best single WVC model. From these results, it appears that the BMA approach is particularly useful for longer lead times.

3. The proposed algorithm combining the BMA technique with the WVC model helped in combining model results from the multiple wavelet model setups and thereby helped in improving the model performance overall. By using the BMA based multiple wavelet ensemble WVC model, different flow regimes were captured properly. Also, it was observed that the weights from the BMA based WVC model provided insights into the type of multi-wavelet combination model that is suitable for a specific time scale. For example, the Haar and db2 wavelets were found to be suitable for daily forecasting, whereas for monthly forecasting, a combination of Haar, db2 and sym2 wavelets provided better forecasts. In the case of weekly forecasting, the model using db2, db4 and Haar wavelets performed the best.

Recommended future research stemming from this study could include:

1. Exploring multi-step ahead forecasting (for example 1, 2, and 3 day ahead; 1, 2, and 3 week ahead; and 1, 2 and 3 month ahead forecasting) using the newly proposed BMA based ensemble multi-wavelet Volterra model.
2. Using different weighting strategies for different flow regimes.
3. Applying the method in other study areas and for different applications (for example groundwater level forecasting, water demand forecasting, etc.) to assess the new BMA WVC approach for different hydrological forecasting applications.

**Table 5**  
Summary results for the different models for Stations I and II at a weekly scale.

Performance measures	Best single wavelet Volterra (WVC)	Average WVC ensemble	BMA based WVC ensemble
<i>Station I</i>			
RMSE (m <sup>3</sup> /s)	52.14	54.5	46.58
MAE (m <sup>3</sup> /s)	26.43	26.43	25.01
Correlation coefficient	0.901	0.878	0.92
NSC	0.79	0.75	0.81
<i>Station II</i>			
RMSE (m <sup>3</sup> /s)	33.18	33.35	26.45
MAE (m <sup>3</sup> /s)	16.07	18.28	13.44
Correlation coefficient	0.89	0.87	0.950
NSC	0.81	0.80	0.87

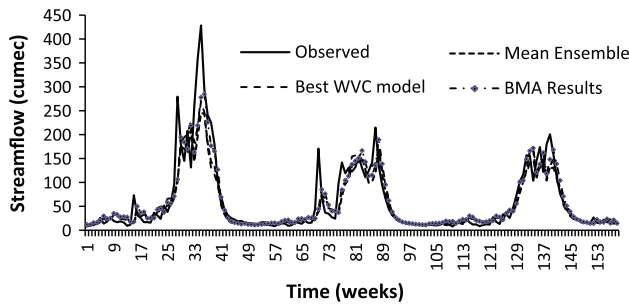


Fig. 12a. Validation results for different models at Station I.

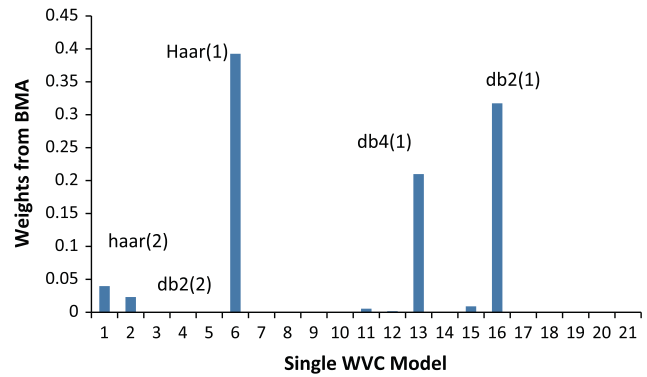


Fig. 13. A single set of BMA weights computed over the entire training period for the BMA WVC models at a daily scale at Station I.

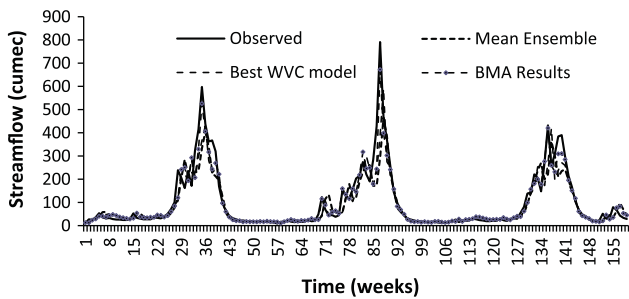


Fig. 12b. Validation results for different models at Station II.

4. Exploring multivariate forecasting (so in addition to flow, including other variables such as precipitation and temperature as input variables).
5. In this study we used existing wavelets for extracting the features from the stream flow time series, however, designing an adaptive wavelet transform matching the characteristics and features of the times series under study could potentially improve forecasting performance significantly.

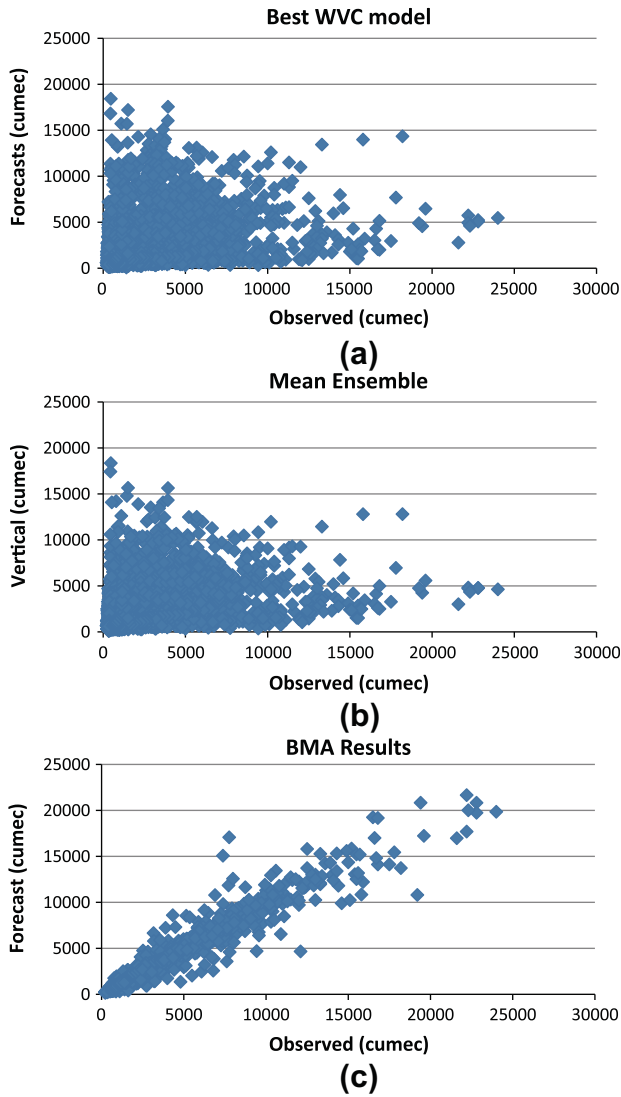
**Table 6**  
Different single wavelet Volterra models for one step ahead forecasting at a daily scale.

Wavelet used	Level of decomposition	Memory (m) at each level {DW's, C}	RMSE (m <sup>3</sup> /s)	MAE (m <sup>3</sup> /s)	NSC
<i>Station I</i>					
Haar	1	{2,2}	572.11	217.56	0.96
db2	1	{2,2}	790.48	320.12	0.93
db2	1	{3,2}	817.02	330.25	0.917
db2	2	{2,1,2}	891.78	393.28	0.90
db2	3	{1,2,2,1}	1063.7	479.99	0.86
Haar	2	{1,1,1,2}	1300.56	584.74	0.79
Haar	3	{2,2,2,1}	1404.78	700.78	0.72
Sym4	3	{2,2,2,2}	2000.23	1208.3	0.35
Sym5	2	{1,2,2}	2088.63	1131.27	0.45
db4	2	{2,2,2}	1299.89	626.40	0.79
db4	2	{2,2,1}	1331.3	658.47	0.77
db5	1	{2,2}	1071.23	471.87	0.85
db4	1	{3,2}	1006.16	433.32	0.88
db4	3	{2,2,2,2}	1507.87	768.63	0.71
<i>Station II</i>					
Haar	1	{2,2}	669.46	292.91	0.98
db2	1	{2,2}	932.00	418.33	0.96
db2	2	{2,3,1}	1140.5	520.78	0.95
db4	1	{2,1}	1154.03	540.78	0.94
Sym4	1	{2,1}	1825.55	900.68	0.86
Sym5	1	{2,2}	1922.17	1000.7	0.81



**Table 7**  
Summary of results for the different models for Stations I and II at a daily scale.

Performance measures	Best single wavelet Volterra (WVC)	Average WVC ensemble	BMA based WVC ensemble
<i>Station I</i>			
RMSE (m <sup>3</sup> /s)	572.11	582.7	512.45
MAE (m <sup>3</sup> /s)	217.56	240.5	200.96
Correlation Coefficient	0.98	0.96	0.99
NSC	0.96	0.945	0.98
<i>Station II</i>			
RMSE (m <sup>3</sup> /s)	669.46	650.79	600.78
MAE (m <sup>3</sup> /s)	292.91	276.97	260.21
Correlation Coefficient	0.99	0.99	0.992
NSC	0.98	0.985	0.99



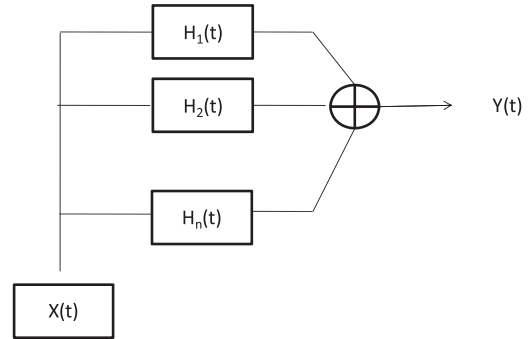
**Fig. 14.** Scatter plot for one day ahead forecasts at Station I using (a) single best WVC model, (b) mean ensemble WVC model and (c) BMA based WVC model.

**Acknowledgements**

This research was partially funded by an NSERC Discovery Grant held by Jan Adamowski.

**Appendix A. Volterra Model**

Volterra series of integral operators was introduced by Volterra (1930) and has since developed into a powerful generic nonlinear



**Fig. A.1.** n<sup>th</sup> order Volterra Representation.

modelling tool. The domain of applied hydrology has also seen application of Volterra representation based approach and some notable contributions have come from Amorocho and Orlob (1961) and Amorocho and Brandstetter, (1971).

$$\begin{aligned}
 y(t) = & \int_0^t h_1(\tau_1)X(\tau - \tau_1)d\tau_1 + \int_0^t \int_0^t h_2(\tau_1, \tau_2)X(\tau - \tau_1) \\
 & X(\tau - \tau_2)d\tau_1d\tau_2 + \int_0^t \int_0^t \int_0^t h_2(\tau_1, \tau_2, \tau_3)X(\tau - \tau_1)X(\tau - \tau_2) \\
 & X(\tau - \tau_3)d\tau_1d\tau_2d\tau_3 + \dots \dots \dots \quad (A.1)
 \end{aligned}$$

Volterra representation is a simple extension of the Taylor series expansion for nonlinear autonomous causal systems with memory and may be written as

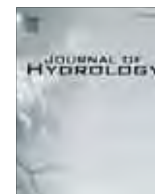
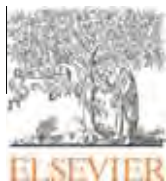
$$y(t) = H_1[x(t)] + H_2[x(t)] + H_3[x(t)] + \dots \dots H_n[x(t)] + \dots \quad (A.2)$$

In the preceding equation, Y(t) represents the system output and x(t) denotes the input to the system. The functions  $h_n(\tau_1, \tau_2, \dots, \tau_n)$  are called the Volterra kernels of the system and the transformation  $H_n[x(t)]$  represents a special type of a convolution integral known as the n<sup>th</sup> order Volterra operator. Figure A.1 shows a schematic representation of the Volterra model in the form of a block diagram.

**References**

Adamowski, K., Prokoph, A., Adamowski, J., 2009. Development of a new method of wavelet aided trend detection and estimation. *Hydrol. Process.* 23 (18), 2686–2696.  
 Adamowski, J., Karapataki, C., 2010. Comparison of multivariate regression and artificial neural networks for peak urban water demand forecasting: evaluation of different ANN learning algorithms. *J. Hydrol. Eng.* 15 (10), 729–743.  
 Adamowski, J., Sun, K., 2010. Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *J. Hydrol.* 390 (1–2), 85–91.  
 Adamowski, J., Fung Chan, H., 2012. A wavelet neural network conjunction model for groundwater level forecasting. *J. Hydrol.* 407 (1–4), 28–40.

- Ajami, N.K., Duan, Q., Sorooshian, S., 2007. An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resour. Res.* 43, W01403. <http://dx.doi.org/10.1029/2005WR004745>.
- Amoroch, J., 1973. Nonlinear hydrologic analysis. *Adv. Hydrosci.* 9, 203–251.
- Amoroch, J., Brandstetter, A., 1971. Determination of Nonlinear Functional Response Functions in Rainfall-Runoff Processes. *Water Resour. Res.* 7 (5), 1087–1101.
- Amoroch, J., Orlob, G.T., 1961. Non-linear analysis of hydrologic systems, Water Resources Center. Univ. California (Berkeley) 40, 147.
- Boucher, M.A., Perreault, L., Anctil, F., 2009. Tools for the assessment of hydrological ensemble forecasts obtained by neural networks. *J. Hydroinform.* 11 (3–4), 297–307.
- Chen, S., Billings, S.A., Luo, W., 1989. Orthogonal least squares methods and their application to non-linear system identification. *Int. J. Control* 50 (5), 1873–1896.
- Chen, G.Y., Xie, W.F., 2007. Pattern recognition with SVM and dual-tree complex wavelets. *Image and Vision Computing* 25 (6), 960–966.
- Chou, C., 2007. Efficient nonlinear modeling of rainfall-runoff process using wavelet compression. *J. Hydrol.* 332 (3–4), 442–455.
- Diskin, M.H., Boneh, A., 1972. Properties of the kernels for time invariant, initially relaxed, second order, surface runoff systems. *J. Hydrol* 17 (1), 115–141.
- Dhanya, Kumar D., 2011. Predictive uncertainty of chaotic daily streamflow using ensemble wavelet networks approach. *Water Resour. Res.*, W06507.
- Duan, Q., Ajami, N.K., Gao, X., Sorooshian, S., 2007. Multimodel ensemble hydrologic prediction using Bayesian model averaging. *Adv. Water Resour.* 30, 1371–1386.
- Fernandez, C., Ley, E., Steel, M.F.J., 2001. Benchmark priors for Bayesian model averaging. *J. Econometrics* 100, 381–427.
- Fundel, F., Zappa, M., 2011. Hydrological ensemble forecasting in mesoscale catchments: Sensitivity to initial conditions and value of reforecasts. *Water Resour. Res.* 47, W09520.
- Georgakakos, K.P., Seo, D.J., Gupta, H., Schake, J., Butts, M.B., 2004. Characterizing streamflow simulation uncertainty through multimodel ensembles. *J. Hydrol.* 298, 222–241.
- Hamil, Thomas, M., 2001. Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Mon. Wea. Rev.* 129, 550–560.
- Hendry, D.F., Clements, M.P., 2002. Pooling of forecasts. *Econometrics J.* 5, 1–26.
- Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial. *Stat. Sci.* 14 (4), 382–417.
- Kallache, M., Rust, H.W., Kropp, J., 2005. Trend assessment: Applications for hydrology and climate research. *Nonlinear Processes Geophys.* 12, 201–210.
- Kisi, Ö., 2009. Wavelet regression model as an alternative to neural networks for monthly streamflow forecasting. *Hydrol. Process.* 23 (25), 3583–3597.
- Kisi, O., 2011. Wavelet regression model as an alternative to neural networks for river stage forecasting. *Water Resour. Manage.* 25, 579–600.
- Labat, D., 2005. Recent advances in wavelet analyses: Part 1. A review of concepts. *J. Hydrol.* 314 (1–4), 275–288.
- Labat, D., Ababou, R., 2001. Introduction of wavelet analyses to rainfall/runoffs relationship for a karstic basin: the case of Liq-Atherey karstic system (France). *Ground Water* 39 (4), 605–615.
- Labat, D., Ababou, R., Mangin, A., 2000. Rainfall–runoff relations for karstic springs. Part II: continuous wavelet and discrete orthogonal multiresolution analyses. *J. Hydrol.* 238 (3), 149–178.
- Leamer, E.E., 1978. *Specification Searches: Ad hoc Inference with Nonexperimental Data*. Wiley, New York.
- Ley, E., Steel, M.F.J., 2009. On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. *J. Appl. Econometrics* 24, 651–674.
- Maheswaran, R., Khosa, R., 2011a. Multiscale nonlinear model for monthly stream flow forecasting: a wavelet based approach. *J. Hydroinform.* 14 (2), 424–442.
- Maheswaran, R., Khosa, R., 2011b. Comparative study of different wavelets for hydrologic forecasting. *Comput. Geosci.* 46, 284–295.
- Maheswaran, R., Khosa, R., 2011c. Wavelet-Volterra coupled model for monthly stream flow forecasting. *J. Hydrol.* 450–451, 320–335.
- Min, C., Zellner, A., 1993. Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. *J. Econometrics* 56 (1–2), 89–118.
- Nalley, D., Adamowski, J., Khalil, B., 2012. Using discrete wavelet transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954–2008). *J. Hydrol.* 475, 204–228.
- Neupauer, R.M., Powell, K.L., Qi, X., Lee, D.H., Villhauer, D.A., 2006. Characterization of permeability anisotropy using wavelet analysis. *Water Resour. Res.* 42, W07419.
- Nourani, V., Alami, M.T., Aminfar, M.H., 2008. A combined neural-wavelet model for prediction of Ligvancai watershed precipitation. *Eng. Appl. Artif. Intel.* 22 (3), 466–472.
- Nourani, V., Komasi, M., Mano, A., 2009. A multivariate ANN-wavelet approach for rainfall-runoff modeling. *Water Resour. Manage.* 23 (14), 2877–2894.
- Parrish, M.A., Moradkhani, H., DeChant, C.M., 2012. Toward reduction of model uncertainty: integration of Bayesian model averaging and data assimilation. *Water Resour. Res.* 48, 1–18.
- Partal, T., 2009. Modelling evapotranspiration using discrete wavelet transform and neural networks. *Hydrol. Process.* 23, 3545–3555.
- Partal, T., Küçük, M., 2006. Long-term trend analysis using discrete wavelet components of annual precipitations measurements in Marmara region Turkey. *Phys. Chem. Earth* 31 (18), 1189–1200.
- Raftery, A.E., Madigan, D., Hoeting, J., 1997. Bayesian model averaging for linear regression models. *J. Am. Stat. Assoc.* 92, 179–191.
- Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Weather Rev.* 133, 1155–1174.
- Renaud, O., Starck, J., Murtagh, F., 2005. Wavelet based combined signal filtering and prediction. *IEEE Trans. Syst. Man Cybern. B Cybern.* 35 (6), 1241–1251.
- Rings, J., Vrugt, J.A., Schoups, G., Huisman, J.A., Vereecken, H., 2012. Bayesian model averaging using particle filtering and Gaussian mixture modeling: theory, concepts, and simulation experiments. *Water Resour. Res.* 48, W05520.
- Sang, Y.F., Wang, D., Wu, J.C., Zhu, Q.P., 2010. Wavelet cross-correlation method for hydrologic time series analysis. *J. Hydraul. Eng.* 41 (11), 1172–1179.
- Sang, Y.F., Wang, Z.G., Liu, C.M., 2012. Discrete wavelet-based trend identification in hydrologic time series. *Hydrol. Process.*
- Shamseldin, A.Y., O'Connor, K.M., 1999. A real-time combination method for the outputs of different rainfall-runoff models. *Hydrological Sciences Journal* 44 (6), 895–912.
- Shensa, M.J., 1992. The discrete wavelet transform: wedding the a trous and Mallat algorithms. *Signal Process.*, *IEEE Trans.* 40 (10), 2464–2482.
- Steel, M.F., 2011. Bayesian model averaging and forecasting. *Bulletin of EU and US Inflation and Macroeconomic Analysis*, 200.
- Stock, J.H., Watson, M.W., 2002. Macroeconomic forecasting using diffusion indexes. *J. Business Econ. Stat.* 20, 147–162.
- Talagrand, O., Vautard, R., Strauss, B., 1999. Evaluation of probabilistic prediction system. *Proc. Workshop on Predictability*, Reading, United Kingdom, ECMWF, pp. 1–25.
- Tiwari, M.K., Chatterjee, C., 2010. Development of an accurate and reliable hourly flood forecasting model using Wavelet-Bootstrap-ANN hybrid approach. *J. Hydrol.* 394, 458–470.
- Vrugt, J.A., Robinson, B.A., 2007. Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. *Water Resour. Res.* 43, W01411.
- Wang, W., Ding, J., 2003. Wavelet network model and its application to the prediction of hydrology. *Science* 1 (1), 67–71.
- Wei, H.L., Billings, S.A., 2004. A unified wavelet-based modelling framework for nonlinear system identification: the WANARX model structure. *Int. J. Control* 77 (4), 351–366.
- Zalachori, I., Ramos, M.H., Garçon, R., Mathevet, T., Gailhard, J., 2012. Statistical processing of forecasts for hydrological ensemble prediction: a comparative study of different bias correction strategies. *Advances in Science & Research*, 8.
- Zhang, X., Zhao, K., 2012. Bayesian Neural Networks for Uncertainty Analysis of Hydrologic Modeling: A Comparison of Two Schemes. *Water Resour. Manage.* 26 (8), 2365–2382.



# Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models



A. Belayneh<sup>a</sup>, J. Adamowski<sup>a,\*</sup>, B. Khalil<sup>a</sup>, B. Ozga-Zielinski<sup>b</sup>

<sup>a</sup> Department of Bioresource Engineering, Faculty of Agricultural and Environmental Sciences, McGill University, 21 111 Lakeshore, Ste Anne de Bellevue, Quebec H9X 3V9, Canada

<sup>b</sup> Department of Environment Protection and Development, Environmental Engineering Faculty, Warsaw University of Technology, ul. Nowowiejska 20, 12 Warsaw 00-653, Poland

## ARTICLE INFO

### Article history:

Received 14 May 2013

Received in revised form 29 October 2013

Accepted 30 October 2013

Available online 7 November 2013

This manuscript was handled by Geoff

Syme, Editor-in-Chief, with the assistance of

Bellie Sivakumar, Associate Editor

### Keywords:

ANN

Support vector regression

SPI

Drought forecasting

Wavelet transforms

Africa

## SUMMARY

Long-term drought forecasts can provide valuable information to help mitigate some of the consequences of drought. Data driven models are suitable forecast tools due to their minimal information requirements and rapid development times. This study compares the effectiveness of five data driven models for forecasting long-term (6 and 12 months lead time) drought conditions in the Awash River Basin of Ethiopia. The Standard Precipitation Index (SPI 12 and SPI 24) was forecasted using a traditional stochastic model (ARIMA) and compared to machine learning techniques such as artificial neural networks (ANNs), and support vector regression (SVR). In addition to these three model types, wavelet transforms were used to pre-process the inputs for ANN and SVR models to form WA-ANN and WA-SVR models; this is the first time that WA-SVR models have been explored and tested for long-term SPI forecasting. The performances of all models were compared using RMSE, MAE,  $R^2$  and a measure of persistence. The forecast results indicate that the coupled wavelet neural network (WA-ANN) models were better than all the other models in this study for forecasting SPI 12 and SPI 24 values over lead times of 6 and 12 months in the Awash River Basin.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

Drought, which occurs when there is a deficit in precipitation compared to the long-term average, has many impacts. Of all the extreme climate induced events, droughts have the most complex consequences, mainly due to the difficulty in identifying their inception and their end (Wilhite, 1993). In addition, droughts impact a wide range of geographic areas. Globally, 22% of the economic damage caused by natural disasters and 33% of the damage in terms of the number of persons affected can be attributed to drought (Keshavarz et al., 2013). The impacts of drought are more severe in sub-Saharan Africa, where rain-fed agriculture comprises 95% of all agriculture. This dependence of rain-fed agriculture leaves sub-Saharan Africa vulnerable to the impacts of drought. For example, in 2009, poor rainfall led to increased droughts and an increase of 53 million food insecure people in the region (Husak et al., 2013).

Due to a slow evolution in time, drought is often a phenomenon whose consequences take a significant amount of time with respect to its inception to be perceived by the socio-economic sector

(Cancelliere et al., 2007b). This feature allows for the mitigation of some of the impacts of drought, provided there is an effective monitoring system to provide information to decision makers. Effective monitoring of droughts can significantly help early warning systems. In sub-Saharan Africa, effective rainfall monitoring contributes to the allocation of aid during periods of drought (Husak et al., 2013). An important aspect of drought monitoring and the development of an early warning system is the ability to effectively forecast future drought events. Forecasting future drought events in a region is very important for finding sustainable solutions to water management and risk assessment of drought occurrences (Bordi and Sutera, 2007).

Drought forecasts can be done using either physical/conceptual or data driven models. While physical/conceptual models are good at providing insight into catchment processes, they have been criticised for being difficult to implement for forecasting applications, requiring many different types of data and resulting in models that are overly complex (Beven, 2006). In contrast, data driven models have minimum information requirements, rapid development times, and have been found to be accurate in various hydrological forecasting applications (Adamowski, 2008).

Data driven stochastic models have traditionally been used for drought forecasting. Autoregressive integrated moving average

\* Corresponding author. Tel.: +1 514 398 7786; fax: +1 514 398 8387.

E-mail address: [jan.adamowski@mcgill.ca](mailto:jan.adamowski@mcgill.ca) (J. Adamowski).

models (ARIMA) (Mishra and Desai, 2005, 2006; Mishra et al., 2007; Han et al., 2010) have been the most widely used stochastic models for hydrologic drought forecasting. Stochastic models are linear models and are limited in their ability to forecast non-linear data. To effectively forecast non-linear data, researchers in the last two decades have increasingly begun to utilize artificial neural networks (ANNs) to forecast hydrological data. ANNs have been used in a number of studies as a drought forecasting tool (Mishra and Desai, 2006; Morid et al., 2007; Bacanli et al., 2008; Barros and Bowden, 2008; Cutore et al., 2009; Karamouz et al., 2009; Marj and Meijerink, 2011; Mishra and Nagarajan, 2012).

Along with ANN models, Support Vector Machines (SVMs) are a machine learning technique that has become increasingly popular in hydrologic forecasting. SVMs can be divided into two main techniques, the Support Vector Classification (SVC) and Support Vector Regression (SVR), which address problems of classification and regression, respectively (Gao et al., 2001). There are several studies where SVR was used in hydrological forecasting. Khan and Coulibaly (2006) found that a SVR model performed better than ANNs in 3–12 month predictions of lake water levels. Kisi and Cimen (2009) used SVRs to estimate daily evaporation. A study by Belayneh and Adamowski (2012) used SVR models to forecast the SPI in the Awash River Basin. The research presented in the current paper complements the Belayneh and Adamowski (2012) study by forecasting the SPI over a larger selection of stations in the same area, and by coupling, for the first time, SVR models with wavelet transforms.

The ability of these machine learning techniques (i.e. ANNs and SVRs) to forecast non-stationary data is limited. To overcome this limitation, wavelet analysis has begun to be explored in hydrologic forecasting. Wavelet transforms are often used as a data pre-processing tool in order to reveal aspects of a time series that other signal processing techniques cannot. Wavelet analysis has been applied to examine rainfall-runoff relationships in Karstic watersheds (Labat et al., 1999), to evaluate rainfall-runoff models (Lane, 2007), to forecast river flow (Adamowski, 2008; Adamowski and Sun, 2010), to forecast groundwater levels (Adamowski and Chan, 2011), to forecast urban water demand (Adamowski et al., 2012) and for the purposes of drought forecasting (Kim and Valdes, 2003; Ozger et al., 2012; Mishra and Singh, 2012; Belayneh and Adamowski, 2012).

The main objective of the current study is to compare the effectiveness of machine learning methods (in this case the ANN and SVR methods) as well as machine learning methods pre-processed with wavelet transforms (WA-ANN and WA-SVR models), for long-term drought forecasting in arid regions (in this case the Awash River Basin of Ethiopia). All these methods were compared to a traditional stochastic method (ARIMA model), which was used for comparative purposes (to compare the performance of the newer methods with a widely used traditional method). The standardized precipitation index (SPI), a meteorological drought index, was the drought index forecasted in this study, as it is a good indicator of the variability of East African droughts (Ntale and Gan, 2003). SPI 12 and SPI 24 were forecast for lead times of 6 and 12 months; SPI 12 and SPI 24 are good indicators of long-term drought conditions. The coupling of wavelet transforms with SVR models for the purpose of forecasting the SPI has not been explored to date in the literature.

## 2. The Standardized Precipitation Index

The Standardized Precipitation Index (SPI) was developed by McKee et al. (1993). The SPI is a meteorological drought index as it is based solely on precipitation data. Two main advantages arise from the use of the SPI index. First, as the index is based on

precipitation alone its evaluation is relatively easy (Cacciamani et al., 2007). Second, the index makes it possible to describe drought on multiple time scales (Tsakiris and Vangelis, 2004; Mishra and Desai, 2006; Cacciamani et al., 2007). The SPI can only be computed when a sufficiently long (at least 30 years) and possibly continuous time-series of monthly precipitation data is available (Cacciamani et al., 2007). SPI calculation begins by selecting a suitable probability density function to describe the precipitation data (Cacciamani et al., 2007). The cumulative probability of an observed precipitation amount is computed after an appropriate density function is chosen. The inverse normal (Gaussian) function is then applied to the probability (Cancelliere et al., 2007a). For each rainfall gauge in this study the gamma distribution function was selected to fit the rainfall data. Alternatively, a lognormal or an exponential distribution can be used to model the precipitation. In this paper, we followed Edossa et al. (2010) on selecting the gamma distribution function to fit the rainfall data. Detailed computation of the SPI can be found in Cacciamani et al. (2007).

The SPI is a z-score and represents an event departure from the mean, expressed in standard deviation units. The SPI is a normalized index in time and space and this feature allows for the comparison of SPI values among different locations. SPI values can be categorized according to classes (Cacciamani et al., 2007). Normal conditions are established from the aggregation of two classes:  $-1 < \text{SPI} < 0$  (mild drought) and  $0 \leq \text{SPI} \leq 1$  (slightly wet). SPI values are positive or negative for greater or less than mean precipitation, respectively. Variance from the mean is a probability indication of the severity of the flood or drought that can be used for risk assessment (Morid et al., 2006). The more negative the SPI value for a given location, the more severe the drought. The time series of the SPI can be used for drought monitoring by setting application-specific thresholds of the SPI for defining drought beginning and ending times. Accumulated values of the SPI can be used to analyze drought severity. In this study, the SPI\_SL\_6 program developed by the National Drought Mitigation Centre, University of Nebraska-Lincoln, was used to compute time series of drought indices (SPI) for each station in the basin and for each month of the year at different time scales. A drought event occurs at the time when the value of the SPI is continuously negative; the event ends when the SPI becomes positive. Even though the SPI is a meteorological drought index, it can be used to interpret different aspects of drought. This study only forecast SPI 12 and SPI 24, which are tied to long-term drought conditions, which is the focus of this study. This study was interested in providing information about drought conditions that affect streamflow, groundwater or other hydrological systems within the Awash River basin. Shorter SPI runs, such as SPI 3 and SPI 6, are representative of agricultural drought conditions, which are measured by a deficit in soil moisture content (Mishra and Desai, 2006).

## 3. Awash River Basin

In this study, the SPI was forecast in the Awash River Basin of Ethiopia (Fig. 1). Forecasts were made and compared for twelve stations within the basin. Drought is a common occurrence in the Awash River Basin (Edossa et al., 2010), and with approximately 90% of the population engaged in agricultural activities the area is especially vulnerable to the effects of drought (Desalegn et al., 2006). The heavy dependence of the population on rain-fed agriculture has made the people and the country's economy extremely vulnerable to the impacts of droughts. Current monthly and seasonal drought forecasts are done using the normalized difference vegetation index (NDVI). While the NDVI is an effective drought index, it is sensitive to changes in vegetation and has limitations in areas where vegetation is minimal. Forecasts of SPI



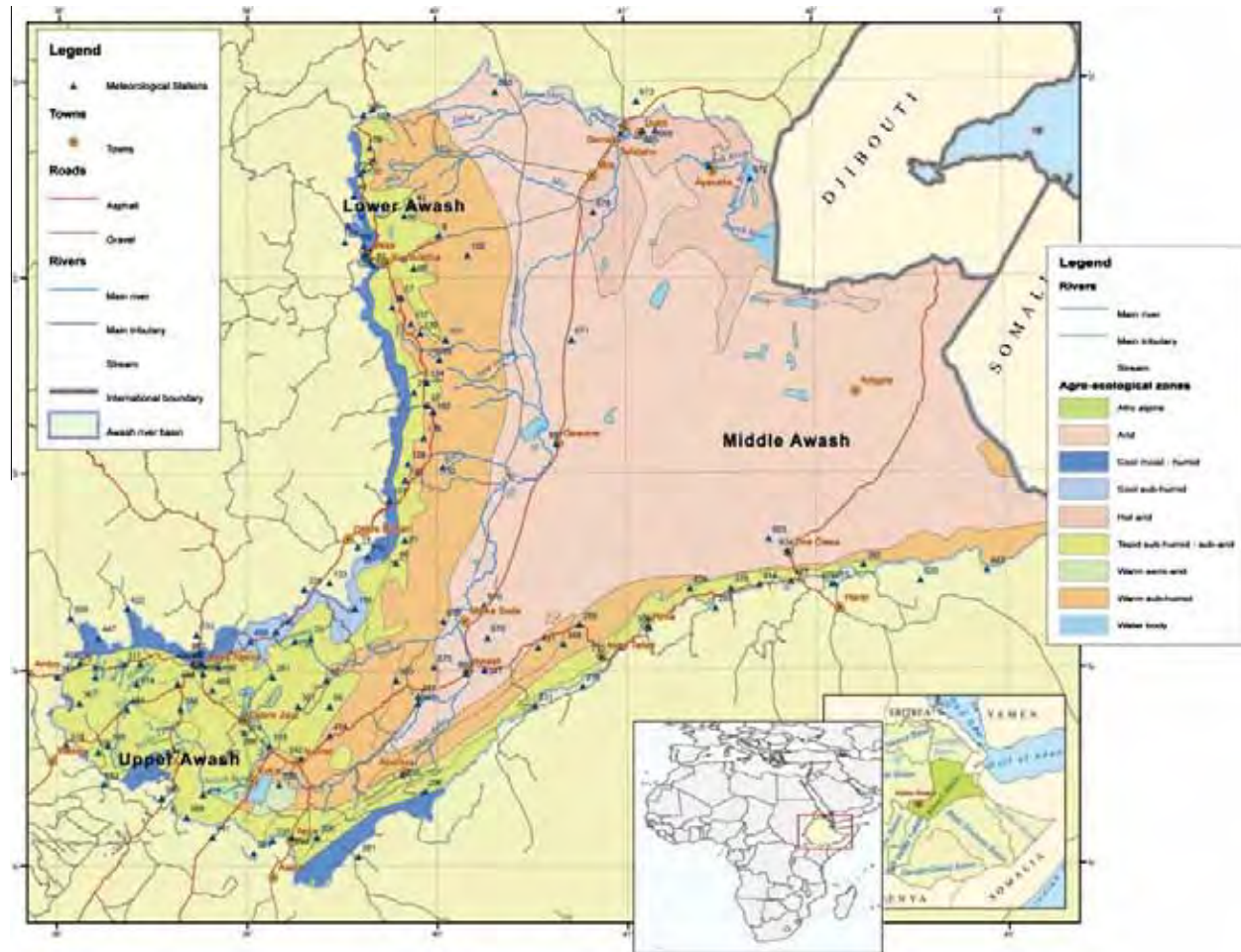


Fig. 1. The Awash River Basin. Ministry of Water Resources, Ethiopia. Agricultural Water Management Information System. [http://www.mowr.gov.et/AWMISSET/images/Awash\\_agroecologyv3.pdf](http://www.mowr.gov.et/AWMISSET/images/Awash_agroecologyv3.pdf) Accessed 06-June-2013.

in this case SPI 12 and SPI 24) are not dependent on vegetative cover and can be used as another tool for drought forecasts within the basin and the country as a whole to complement the NDVI forecasts.

The mean annual rainfall of the basin varies from about 1600 mm in the highlands to 160 mm in the northern point of the basin. The total amount of rainfall also varies greatly from year to year, resulting in severe droughts in some years and flooding in others. The total annual surface runoff in the Awash Basin amounts to some  $4900 \times 10^6 \text{ m}^3$  (Edossa et al., 2010). Effective forecasts of the SPI can be used for mitigating the impacts of hydrological drought that manifests as a result of rainfall shortages in the area. The climate of the Awash River Basin varies between a mild temperate climate in the Upper Awash sub-basin to a hot semi-arid climate in both the Middle and Lower sub-basins. The Awash River Basin supports farming, from the growth of lowland crops such as maize and sesame to pastoral farming practices. Rainfall records from 1970 to 2005 were used to generate SPI 12 and SPI 24 time series from each station. Descriptive statistics for precipitation at the rainfall stations is shown in Table 1.

The normal ratio method, recommended by Linsley et al. (1988), was used to estimate the missing rainfall records at any stations that had incomplete precipitation records. With this method, rain depths for missing data are estimated from observations at three stations as close to, and as evenly spaced around the station with missing records, as possible. The distance matrix was

established for all rain gauge stations in the basin based on their geographic locations in order to assess the proximity of stations with each other. All data sets were normalized using:

$$X_n = \frac{X_0 - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

where  $X_0$  and  $X_n$  represent the original and normalized data respectively, while  $X_{\min}$  and  $X_{\max}$  represent the minimum and maximum value among the original data.

## 4. Model development

### 4.1. ARIMA models

ARIMA models were developed based on the Box and Jenkins approach and consist of three steps: model identification, parameterization and validation. The general non-seasonal ARIMA model may be written as (Box and Jenkins, 1976):

$$z_t = \frac{\theta(B)a_t}{\phi(B)\nabla^d} \quad (2)$$

$$\phi B = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \quad (3)$$

and

**Table 1**  
Descriptive Statistics of the Awash River Basin.

Basin	Station	Mean annual precipitation (mm)	Max annual precipitation (mm)	Standard deviation (mm)
Upper Awash Basin	Bantu	91	647	111
	Liben			
	Ginchi	97	376	90
	Sebeta	111	1566	172
	Ejersalele	67	355	75
	Ziquala	100	583	110
Middle Awash Basin	Modjo	76	542	92
	Wolenchiti	76	836	95
	Gelemsso	77	448	75
	Dire Dawa	51	267	54
Lower Awash Basin	Eliwuha	44	374	57
	Dubti	87	449	89
	Bati	26	268	40

$$\theta B = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \tag{4}$$

where  $z_t$  is the observed time series and  $B$  is a back shift operator.  $\phi(B)$  and  $\theta(B)$  are polynomials of order  $p$  and  $q$ , respectively. The orders  $p$  and  $q$  are the order of non-seasonal auto-regression and the order of non-seasonal moving average, respectively. Random errors  $a_t$  are assumed to be independently and identically distributed with a mean of zero and a constant variance.  $\nabla^d$  describes the differencing operation to data series to make the data series stationary and  $d$  is the number of regular differencing.

The first step in developing ARIMA models was determining the stationarity of a time series. To determine stationarity, the NUMxl Excel add-on was used. Once non-stationarity is removed the autocorrelation (ACF) and partial autocorrelation functions (PACF) were used to determine the correlation structure of the data. Once the significant lags were determined using the ACF and PACF, different combinations were used to determine the optimal model structure. For instance, if the ACF and PACF for SPI 12 showed significant lags at 5 and 2 lags respectively, different combinations for  $p$  and  $q$  were used with intervals between 1 and 5 for  $p$  and 1 and 2 for  $q$ . The selection of ARIMA models was determined based on both the accuracy and precision of models. The combinations that provided the most accurate forecast models as measured by the MAE and the most precise models as measured by the RMSE were chosen. The details on the development of ARIMA models for SPI time series can be found in the works of Mishra and Desai (2005) and Mishra et al. (2007). For all ARIMA models, the data was partitioned so that 90% of the data was a calibration set and 10% of the data was a validation set.

#### 4.2. ANN models

The advantage of using ANNs is their parsimonious data requirements, rapid execution time and ability to produce models where the relationship between inputs and outputs are not fully understood. The ANN models used in this study have a feed forward Multi-layer perceptron (MLP) architecture which was trained with the Levenberg–Marquardt (LM) back propagation algorithm. MLPs have often been used in hydrologic forecasting due to their simplicity. MLPs consist of an input layer, one or more hidden layers, and an output layer (Kim and Valdes, 2003):

$$y'_k(t) = f_0 \left[ \sum_{j=1}^m w_{kj} \cdot f_n \left( \sum_{i=1}^N w_{ji} x_i(t) + (w_{j0}) + w_{k0} \right) \right] \tag{5}$$

where  $N$  is the number of samples,  $m$  is the number of hidden neurons,  $x_i(t)$  = the  $i$ th input variable at time step  $t$ ;  $w_{ji}$  = weight that

connects the  $i$ th neuron in the input layer and the  $j$ th neuron in the hidden layer;  $w_{j0}$  = bias for the  $j$ th hidden neuron;  $f_n$  = activation function of the hidden neuron;  $w_{kj}$  = weight that connects the  $j$ th neuron in the hidden layer and  $k$ th neuron in the output layer;  $w_{k0}$  = bias for the  $k$ th output neuron;  $f_0$  = activation function for the output neuron; and  $y'_k(t)$  is the forecasted  $k$ th output at time step  $t$  (Kim and Valdes, 2003).

For ANN model development, the determination of the architecture of the model is very important. The optimal number of neurons in the input layer was determined by trial and error. As the SPI requires only precipitation for its computation it was the only input used. The SPI was lagged to generate several neurons in the input layer and the number of neurons that provided the lowest RMSE values was chosen as the appropriate number. Traditionally, the number of hidden neurons for ANN models is selected via a trial and error method. However a study by Wanas et al. (1998) empirically determined that the best performance of a neural network occurs when the number of hidden neurons is equal to  $\log(N)$ , where  $N$  is the number of training samples. Another study conducted by Mishra and Desai (2006) determined that the optimal number of hidden neurons is  $2n + 1$ , where  $n$  is the number of input neurons. In this study, the optimal number of hidden neurons was determined to be between  $\log(N)$  and  $(2n + 1)$ . For example, if using the method proposed by Wanas et al. (1998) gave a result of 4 hidden neurons and using the method proposed by Mishra and Desai (2006) gave 7 hidden neurons, the optimal number of hidden neurons was assumed to be between 4 and 7; thereafter the optimal number was chosen via trial and error. These two methods helped establish an upper and lower bound for the number of hidden neurons.

The ANN models used to forecast the SPI were recursive models. A recursive ANN model is similar to an ARIMA model in terms of the forecasting approach, and forecasts one time step ahead. For subsequent forecasts the network is applied recursively, using the previous predictions as inputs. Recursive ANN models have been shown to be effective for forecasts of the SPI for long lead times (e.g., Mishra and Nagarajan, 2012). In addition, preliminary analysis conducted for each station in our study indicated that recursive models were more accurate than direct multistep models, and as such the recursive forecasting approach was used in this study. For all the ANN models, 80% of the data was used to train the models, while the remaining 20% of the data was divided into a testing and validation set with each set comprising 10% of the data.

#### 4.3. SVR models

SVR models adhere to the structural risk minimization principle as opposed to the empirical risk minimization principle used by conventional neural networks (Vapnik, 1995). As a result, these models reduce the generalization error as opposed to the training error. With SVR, the purpose is to estimate a functional dependency  $f(\vec{x})$  between a set of sampled points  $X = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\}$  taken from  $R^n$  and target values  $Y = \{y_1, y_2, \dots, y_n\}$  with  $y_i \in R$  (the input and target vectors ( $x_i$ 's and  $y_i$ 's) refer to the monthly records of the SPI index). Detailed descriptions of SVR model development can be found in Cimen (2008).

All SVR models were created using the OnlineSVR software created by Parrella (2007), which can be used to build support vector machines for regression. The data was partitioned into two sets: a calibration set and a validation set. 90% of the data was partitioned into the calibration set while the final 10% of the data was used as the validation set. Unlike neural networks the data can only be partitioned into two sets with the calibration set being equivalent to the training and testing sets found in neural networks. All inputs and outputs were standardized between 0 and 1.

All SVR models used the non-linear radial basis function (RBF) kernel. As a result, each SVR model consisted of three parameters that were selected: gamma ( $\gamma$ ), cost ( $C$ ), and epsilon ( $\epsilon$ ). The  $\gamma$  parameter is a constant that reduces the model space and controls the complexity of the solution, while  $C$  is a positive constant that is a capacity control parameter, and  $\epsilon$  is the loss function that describes the regression vector without all the input data (Kisi and Cimen, 2011). These three parameters were selected based on a trial and error procedure. The combination of parameters that produced the lowest RMSE and MAE values for the calibration data sets were selected.

#### 4.4. Wavelet decomposition

The wavelet transform is a mathematical tool that provides a time–frequency representation of a signal in the time domain (Partal and Kisi, 2007). In addition, wavelet analysis can often compress or de-noise a signal (Kim and Valdes, 2003) and thus, is an effective method for dealing with local discontinuities in a given time series. The continuous wavelet transform (CWT) of a signal  $x(t)$  is defined as (Nason and Von Sachs, 1999):

$$W(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} x(t) \psi^* \left( \frac{t - \tau}{s} \right) dt \quad (6)$$

where  $s$  is the scale parameter;  $\tau$  is the translation,  $\psi$  is the mother wavelet and  $*$  corresponds to the complex conjugate (Kim and Valdes, 2003).

CWT is not often used for forecasting due to its complexity and long computation times. Instead, the wavelet is discretized in forecasting applications to simplify the numerical calculations. The discrete wavelet transform (DWT) requires less computation time and is simpler to implement (Nason and Von Sachs, 1999):

$$\psi_{j,k}(t) = \frac{1}{\sqrt{|s_0^j|}} \psi \left( \frac{t - k\tau_0 s_0^j}{s_0^j} \right) \quad (7)$$

where  $j$  and  $k$  are integers that control the scale and translation respectively, while  $s_0 > 1$  is a fixed dilation step (Cannas et al., 2006) and  $\tau_0$  is a translation factor that depends on the aforementioned dilation step.

One of the inherent limitations of using the DWT for forecasting applications is that it is not shift invariant (i.e. if we change values at the beginning of our time series, all of the wavelet coefficients will change). To overcome this problem, a redundant algorithm, known as the à trous algorithm, can be used and is given by (Mallat, 1998):

$$C_{i+1}(k) = \sum_{l=-\infty}^{+\infty} h(l) c_i(k + 2^i l) \quad (8)$$

where  $h$  is the low pass filter and  $C_{i+1}(k)$  is the original time series. To extract the details,  $w_i(k)$ , that were eliminated in Eq. (8), the smoothed version of the signal is subtracted from the coarser signal that preceded it, given by (Murtagh et al., 2003):

$$w_i(k) = c_{i-1}(k) - c_i(k) \quad (9)$$

where  $c_i(k)$  is the approximation of the signal and  $c_{i-1}(k)$  is the coarser signal. Each application of Eqs. (8) and (9) results in a smoother approximation and extracts a higher level of detail. Finally, the non-symmetric Haar wavelet can be used as the low pass filter for the à trous algorithm to prevent any future information from being used during the decomposition (Renaud et al., 2002).

The selection of the appropriate wavelet transform for an application requires a prior understanding of the attributes of the candidate wavelet. The shift invariant property of the 'à trous' Haar

wavelet makes it the most suitable wavelet for forecasting applications (Maheswaran and Khosa, 2012b), and this is why it was used in this study. Furthermore, a study by Kim and Valdes (2003) found the 'à trous' wavelet to be more effective for forecasting studies than the Morlet or Daubechies wavelets. The aim of the 'à trous' wavelet is to fill any gaps with redundant information that is obtained from the original series. The redundant information provides a basis for enhanced forecasting accuracy (Maheswaran and Khosa, 2012a). The Haar wavelet, which is a low-pass filter, is concentrated over the narrowest support band, and therefore has good localization properties. This attribute makes the Haar wavelet more suitable for dynamic time series where it is effective at detecting changes within the time series (Maheswaran and Khosa, 2012b). The 'à trous' wavelet is appropriate for forecasting due to its localization capability in both the time and frequency domains, and the Haar wavelet as a low-pass filter is appropriate for detecting any changes that may occur within the 35 year SPI record in this study due to its localization properties as a result of its energy being concentrated over a narrow support band.

When conducting wavelet analysis, the number of decomposition levels that is appropriate for the data must be carefully selected. In this study, each SPI time series was decomposed between 1 and 8 levels. Fig. 2 depicts the SPI 12 time series of the Eliwuha station and the approximation series at different decomposition levels. The results were compared at all decomposition levels to determine the appropriate decomposition level. Fig. 2 indicates that the higher the level of decomposition, the less likely the transformed signal represents the original time series. Hence, the transformed times series was chosen after decomposing the original time series to level three or level four. The results from a decomposition of level three or four provided the most accurate forecast results as measured by the performance measures used in this study. Once a given time series was successfully decomposed, it was used as an input for either ANN or SVR models. Instead of using the original SPI data and its subsequent lags, a new wavelet transformed time series was used. After pre-processing, the generation of ANN and SVR models, including data partition was done in exactly the same way as the ANN and SVR models without wavelet transforms.

#### 4.5. Performance measures

The following measures of goodness of fit were used to evaluate the forecast performance of all the aforementioned models:

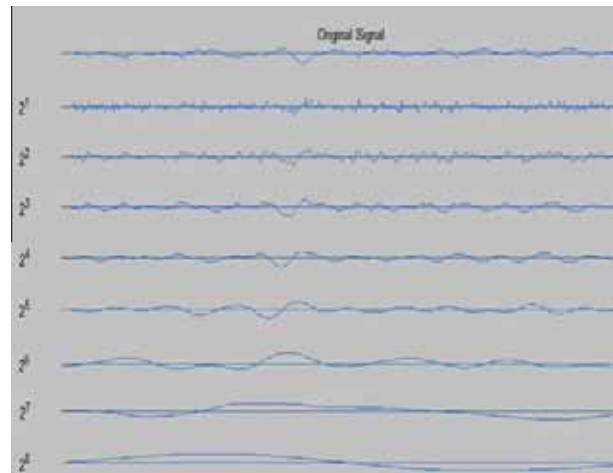


Fig. 2. The SPI 12 time series at the Eliwuha station and the approximation time series at different decomposition levels. The first time series is the original signal followed by decompositions a level 1–8.



$$\text{The coefficient of determination}(R^2) = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (10)$$

$$\text{where } \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i \quad (11)$$

where  $\bar{y}$  is the mean value taken over  $N$ ,  $y_i$  is the observed value,  $\hat{y}_i$  is the forecasted value and  $N$  is the number of data points. The coefficient of determination measures the degree of association among the observed and predicted values.

$$\text{The Root Mean Squared Error(RMSE)} = \sqrt{\frac{SSE}{N}} \quad (12)$$

where  $SSE$  is the sum of squared errors, and  $N$  is the number of samples used.  $SSE$  is given by:

$$SSE = \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (13)$$

with the variables already having been defined.

$$\text{The Mean Absolute Error(MAE)} = \frac{\sum_{i=1}^N |\hat{y}_i - y_i|}{N} \quad (14)$$

The MAE is used to measure how close forecasted values are to the observed values. It is the average of the absolute errors.

The results in this section were also compared via the persistence index ( $PERS$ )

$$PERS = 1 - \frac{SSE}{SSE_{naive}} \quad (15)$$

$$\text{where } SSE_{naive} = \sum (y_i - y_{i-L})^2 \quad (16)$$

As mentioned above,  $SSE$  is the sum of squared errors.  $y_{i-L}$  is the estimate from a persistence model that takes the last observation (at time 1 minus the lead time ( $L$ )) (Tiwari and Chatterjee, 2010). A value of  $PERS$  smaller or equal to 0 indicates that the model under study performs worse or no better than the easy to implement naïve model. A  $PERS$  value of 1 is obtained when the model under study provides exact estimates of observed values.

### 5. Results and discussion

In this study, the proposed forecast models for SPI 12 and SPI 24 are presented for forecast lead times of 6 and 12 months. A SPI 12 forecast of 6 months lead time represents a 6 month warning time for SPI 12, and a 12 month lead time represents a 12 month warning time and shows the variation in precipitation from year to year. Table 2 shows the inputs used for the proposed data driven models at the Eliwuha station (6 months lead time). Table 2 shows how the models are applied recursively to reach the final target of a forecast of 6 months lead time. The table is applicable for the input structure of both SPI 12 and SPI 24 (6 month lead time). The performance results of the proposed models for each station are presented in Table 3 through Table 6. As mentioned earlier, models that have a persistence index between 0 and 1 perform better than a naïve model. All the data driven models had a persistence index greater than 0. ARIMA models had a  $PERS$  of 0.36, ANN models had a  $PERS$  of 0.46, SVR models had a  $PERS$  of 0.41, WA-ANN models had a  $PERS$  of 0.58 and WA-SVR models had a  $PERS$  of 0.55 respectively. The results presented are based on the validation data sets. Table 7 shows the performance results for the training/calibration set for SPI 12 forecasts (6 months lead time) at the Eliwuha station. The table shows that the forecast accuracy shown in the validation sets

**Table 2**

The decompositions of the SPI 12 series at the Eliwuha station with the correlation and variation of these decompositions to the original SPI 12 series.

Decomposition level	$R^2$	RMSE
Level 1	0.8408	0.0742
Level 2	0.8408	0.0742
Level 3	0.9727	0.0412
Level 4	0.9458	0.0490
Level 5	0.8408	0.0742
Level 6	0.7033	0.1068
Level 7	0.6671	0.1300
Level 8	0.6129	0.1446

is consistent with the results shown in the training or calibration sets (see Table 8).

The models were forecast one time step ahead and the subsequent result was used as an input in another model and forecast one time step ahead. The lead time in this study refers to the amount of time between the original time series and the final predicted time series. The final output time series was either 6 or 12 months ahead of the original time series. For each rainfall station, forecasts of SPI 12 and SPI 24 were made for 6 and 12 months lead times. These forecasts were made using the five model types outlined above. The results presented in the following sections are from forecasts of SPI 12 (6 months lead time) at the Eliwuha station. Only these results were presented in detail as it would be very difficult to present all the results from each station in detail. The forecast results for all the stations are presented in Tables 3–6, but these are not presented in detail.

#### 5.1. ARIMA model results

The parameters for the ARIMA models were selected based on the ACF and PACF of the time series in question. Once the significant lags were determined from the ACF and PACF, ARIMA models with different combinations were developed and the model that had the lowest RMSE and MAE value was chosen. Figs. 3a and 3b are ACF and PACF figures for the Eliwuha station. The autocorrelation of the time series is significant for lags greater than 10 in Fig. 3a; however no lag greater than 5 was selected for the purpose of parsimony. For this station a combination of parameters from 1 to 5 were tested for  $p$ , while a value of 1 was tested for  $q$ . The combination that provided the lowest RMSE and MAE values was chosen as the model parameters. The results from the ARIMA models as shown in Tables 3 and 5 are significantly less accurate than the forecast results of the other model types (likely because ARIMA models are linear models).

#### 5.2. ANN models

Fig. 4 provides the ANN model results at the Eliwuha station for SPI 12 at 6 months lead time. Fig. 4 indicates that the model has good generalization ability for SPI 12 and very little time shift error as the extreme events for the predicted values correspond to the extreme values of the observed values. With respective extreme drought or extreme precipitation, the model does underestimate

**Table 3**

Model input structure for the Eliwuha station (6 months lead time).

Recursive order	Input structure	Target
1st step	SPI(t-4), SPI(t-3), SPI(t-2), (SPI t-1), SPI (t)	SPI(t+1)
2nd step	SPI(t-3), SPI(t-2), SPI (t-1), SPI (t), SPI(t+1)	SPI(t+2)
3rd step	SPI(t-2), SPI(t-1), SPI(t), (SPI t+1), SPI (t+2)	SPI(t+3)
4th step	SPI(t-1), SPI(t), SPI (t+1), SPI (t+2), SPI(t+3)	SPI(t+4)
5th step	SPI(t), SPI(t+1), SPI(t+2), (SPI t+3), SPI (t+4)	SPI(t+5)
Test step	SPI(t+1), SPI(t+3), SPI (t+4), SPI (t+4), SPI(t+5)	SPI(t+6)



**Table 4**

The best ARIMA, ANN and SVR models for 6 and 12 month forecasts of SPI 12. Column 3 is the ANN architecture detailing the number of nodes in the input, hidden and output layers respectively. In column 11 the parameters of the SVR models are given.

Basin	Station	ANN models	$R^2$	RMSE	MAE	ARIMA (p,d,q)	$R^2$	RMSE	MAE	SVR ( $\gamma, C, \epsilon$ )	$R^2$	RMSE	MAE
<i>6 Month lead time</i>													
Upper	Bantu Liben	5-4-1	0.7013	0.3737	0.3051	(5,0,0)	0.5912	0.8757	0.7714	0.4, 99, 0.008	0.7343	0.2427	0.2063
	Ginchi	5-4-1	0.7120	0.3834	0.3592	(5,1,1)	0.5391	0.8851	0.7459	0.4, 94, 0.008	0.7028	0.2535	0.1984
	Sebeta	5-4-1	0.6993	0.3784	0.3516	(4,0,0)	0.5827	0.8789	0.7407	0.4, 96, 0.008	0.7320	0.2805	0.2007
	Ejersalele	5-4-1	0.7103	0.3751	0.3458	(3,1,0)	0.5261	0.8846	0.7539	0.5, 90, 0.006	0.7127	0.2997	0.1984
	Ziquala	6-4-1	0.7201	0.3765	0.3271	(4,0,0)	0.5713	0.8623	0.7436	0.3, 89, 0.005	0.7382	0.2690	0.2052
Middle	Modjo	5-4-1	0.7064	0.3673	0.3281	(3,1,10)	0.5329	0.8600	0.7333	0.6, 88, 0.008	0.7326	0.2917	0.2044
	Wolenchiti	6-4-1	0.7139	0.3722	0.3481	(5,0,0)	0.5537	0.8665	0.7388	0.9, 90, 0.008	0.7044	0.2884	0.1987
	Gelemsso	5-4-1	0.7045	0.3619	0.3581	(4,0,0)	0.5572	0.8751	0.7459	0.8, 91, 0.007	0.7150	0.2891	0.2147
	Dire Dawa	5-4-1	0.7123	0.3781	0.3440	(5,0,1)	0.5928	0.8731	0.7443	0.4, 92, 0.005	0.7329	0.2838	0.1885
Lower	Dubti	6-4-1	0.7077	0.3409	0.3103	(5,0,0)	0.5472	0.8928	0.7607	0.8, 93, 0.008	0.7346	0.3054	0.1851
	Eliwuha	5-4-1	0.7565	0.3296	0.3019	(4,0,1)	0.5534	0.8857	0.7798	0.8, 95, 0.008	0.7653	0.3136	0.1953
	Bati	6-4-1	0.7210	0.3434	0.3097	(5,0,0)	0.5237	0.8928	0.7940	0.5, 87, 0.009	0.7547	0.2831	0.2162
<i>12 Month lead time</i>													
Upper	Bantu Liben	5-4-1	0.5129	0.4120	0.3742	(4,0,0)	0.4421	0.9556	0.7577	0.4, 90, 0.01	0.6087	0.3723	0.2582
	Ginchi	6-4-1	0.5346	0.4309	0.3781	(4,0,1)	0.4472	0.9015	0.7343	0.5, 91, 0.007	0.6078	0.3794	0.2352
	Sebeta	5-4-1	0.5456	0.4456	0.3981	(5,0,0)	0.4638	0.9123	0.7295	0.6, 97, 0.003	0.6123	0.3912	0.2807
	Ejersalele	6-4-1	0.5422	0.4021	0.3448	(5,0,0)	0.4537	0.9286	0.7416	0.6, 100, 0.01	0.6078	0.3466	0.2051
	Ziquala	7-4-1	0.5437	0.4358	0.3891	(4,0,1)	0.4682	0.9526	0.7321	0.8, 99, 0.03	0.6234	0.3556	0.2599
Middle	Modjo	6-4-1	0.5273	0.3996	0.3568	(4,1,0)	0.4572	0.9725	0.7755	0.5, 96, 0.008	0.6350	0.3482	0.2034
	Wolenchiti	7-4-1	0.5358	0.4180	0.3689	(3,0,0)	0.4589	0.9545	0.7785	0.8, 96, 0.008	0.6267	0.3114	0.1963
	Gelemsso	5-4-1	0.5441	0.4322	0.3889	(3,0,1)	0.4578	0.9592	0.7752	0.6, 94, 0.009	0.6290	0.3300	0.2115
	Dire Dawa	5-4-1	0.5291	0.3900	0.3770	(3,0,0)	0.4280	0.9678	0.7690	0.5, 96, 0.008	0.6443	0.3466	0.2016
Lower	Dubti	6-4-1	0.5127	0.3976	0.3782	(2,0,2)	0.4627	0.9285	0.7904	0.6, 96, 0.004	0.6239	0.3359	0.2212
	Eliwuha	4-4-1	0.5371	0.4304	0.3623	(5,0,0)	0.4632	0.9204	0.7837	0.8, 94, 0.008	0.6219	0.3087	0.1958
	Bati	7-4-1	0.5230	0.4032	0.3703	(4,0,0)	0.4547	0.9890	0.7837	0.5, 96, 0.008	0.6387	0.3783	0.2426

**Table 5**

The best WA-ANN and WA-SVR models for 6 and 12 month forecasts of SPI 12. Column 3 is the ANN architecture detailing the number of nodes in the input, hidden and output layers respectively. In column 7 the parameters of the SVR models are given.

Basin	Station	WA-ANN	$R^2$	RMSE	MAE	WA-SVR	$R^2$	RMSE	MAE
<i>6 Month lead time</i>									
Upper	Bantu Liben	5-6-1	0.8696	0.2023	0.1933	0.4, 99, 0.008	0.8507	0.2409	0.1934
	Ginchi	5-5-1	0.8276	0.2095	0.1831	0.4, 94, 0.008	0.8144	0.2314	0.2136
	Sebeta	5-5-1	0.8815	0.2013	0.1833	0.4, 96, 0.008	0.8380	0.2343	0.2214
	Ejersalele	7-5-1	0.9090	0.2066	0.1821	0.5, 90, 0.006	0.8363	0.2221	0.2054
	Ziquala	6-5-1	0.8616	0.2057	0.1830	0.3, 89, 0.005	0.8330	0.2413	0.2138
Middle	Modjo	5-4-1	0.8953	0.2083	0.1828	0.6, 88, 0.008	0.8732	0.2245	0.2054
	Wolenchiti	7-4-1	0.9332	0.2015	0.1892	0.9, 90, 0.008	0.8644	0.2518	0.2210
	Gelemsso	6-5-1	0.9204	0.2000	0.1845	0.8, 91, 0.007	0.8726	0.2167	0.2017
	Dire Dawa	7-6-1	0.9129	0.2001	0.1928	0.4, 92, 0.005	0.8968	0.2212	0.2036
Lower	Dubti	4-4-1	0.9231	0.2060	0.1847	0.8, 93, 0.008	0.8640	0.2185	0.2084
	Eliwuha	5-4-1	0.9326	0.2088	0.1804	0.8, 95, 0.008	0.8671	0.2440	0.2213
	Bati	6-4-1	0.9005	0.2183	0.1997	0.5, 87, 0.009	0.8441	0.2467	0.2341
<i>12 Month lead time</i>									
Upper	Bantu Liben	5-6-1	0.8034	0.2235	0.2115	0.4, 90, 0.01	0.7535	0.2484	0.2228
	Ginchi	6-5-1	0.8261	0.2416	0.2128	0.5, 91, 0.007	0.7533	0.2455	0.2030
	Sebeta	5-5-1	0.8049	0.2314	0.2126	0.6, 97, 0.003	0.7148	0.2734	0.2622
	Ejersalele	8-6-1	0.8162	0.2208	0.2128	0.6, 100, 0.01	0.7342	0.2645	0.2406
	Ziquala	7-5-1	0.8300	0.2206	0.2128	0.8, 99, 0.03	0.7604	0.2922	0.2711
Middle	Modjo	6-4-1	0.8292	0.2334	0.2242	0.5, 96, 0.008	0.7643	0.2645	0.2449
	Wolenchiti	8-5-1	0.8046	0.2365	0.2247	0.8, 96, 0.008	0.7843	0.2817	0.2733
	Gelemsso	7-6-1	0.8272	0.2354	0.2216	0.6, 94, 0.009	0.7721	0.2996	0.2717
	Dire Dawa	8-6-1	0.8219	0.2362	0.2117	0.5, 96, 0.008	0.7813	0.2674	0.2241
Lower	Dubti	5-5-1	0.8549	0.2406	0.2052	0.6, 96, 0.004	0.7443	0.2745	0.2620
	Eliwuha	6-4-1	0.8473	0.2719	0.2534	0.8, 94, 0.008	0.7641	0.3012	0.2639
	Bati	8-4-1	0.8437	0.2350	0.2248	0.5, 96, 0.008	0.7134	0.2611	0.2418

some instances where the observed SPI value corresponds to extreme conditions (−2 and 2) respectively. Fig. 5 is a scatter plot of the observed and ANN forecast results for the Eliwuha station. The scatter plot in Fig. 5 shows several points significantly below the trend line indicating a certain level of underestimation in the ANN model results. The proposed ANN model from the Eliwuha station had an  $R^2$  of 0.7564, an RMSE of 0.3296 and an MAE of 0.3019, respectively. The  $R^2$  values of 0.7564 shows a good correlation between observed and predicted results at 6-months lead time.

### 5.3. WA-ANN models

The results of all the proposed WA-ANN models can be found in Table 4 for both forecasts of 6 and 12 months lead time. The forecast results of WA-ANN models are more accurate than ANN models according to the performance measures used. The use of wavelets as a pre-processing tool resulted in more accurate and precise models as seen in Tables 4 and 6. These model results have a higher level of correlation between the observed and predicted time series and the RMSE and MAE values are lower compared to

**Table 6**

The best ARIMA, ANN and SVR models for 6 and 12 month forecasts of SPI 24. Column 3 is the ANN architecture detailing the number of nodes in the input, hidden and output layers respectively. In column 11 the parameters of the SVR models are given.

Basin	Station	ANN models	R <sup>2</sup>	RMSE	MAE	ARIMA (p,d,q)	R <sup>2</sup>	RMSE	MAE	SVR (γ,C,ε)	R <sup>2</sup>	RMSE	MAE
<i>6 Month lead time</i>													
Upper	Bantu	4-4-1	0.7832	0.2775	0.2404	(5,1,0)	0.6486	0.5867	0.5735	0.8, 93, 0.08	0.7754	0.3192	0.2967
	Liben	5-4-1	0.7949	0.3302	0.3238	(4,1,0)	0.6302	0.5853	0.5707	0.8, 94, 0.06	0.7581	0.3027	0.2883
	Ginchi	5-4-1	0.7682	0.3421	0.3325	(5,0,1)	0.6682	0.5844	0.5688	0.9, 93, 0.08	0.7961	0.3338	0.2991
	Sebeta	6-4-1	0.7881	0.3347	0.3131	(5,1,0)	0.6427	0.5832	0.5665	0.8, 93, 0.08	0.7515	0.3546	0.3029
	Ejersalele Ziquala	4-4-1	0.7947	0.3619	0.3421	(5,0,1)	0.6362	0.5822	0.5644	0.5, 99, 0.05	0.7862	0.2627	0.2534
Middle	Modjo	4-4-1	0.7915	0.2836	0.2629	(5,0,0)	0.6526	0.6046	0.5421	0.7, 92, 0.06	0.7637	0.3115	0.2951
	Wolenchiti	4-4-1	0.7844	0.3302	0.3026	(4,0,1)	0.5721	0.5927	0.5721	0.6, 94, 0.08	0.7619	0.3143	0.3065
	Gelemsso	5-4-1	0.7916	0.2262	0.2202	(5,0,0)	0.6676	0.5821	0.5626	0.45, 96, 0.01	0.7738	0.3498	0.3431
	Dire Dawa	5-4-1	0.8042	0.2662	0.2632	(3,1,1)	0.6742	0.5780	0.5581	0.4, 89, 0.03	0.7887	0.2933	0.2788
Lower	Dubti	4-4-1	0.8041	0.3397	0.3187	(5,0,0)	0.6248	0.5819	0.5355	0.4, 94, 0.07	0.7983	0.3141	0.2874
	Eliwuha	5-4-1	0.7965	0.3129	0.3017	(5,0,0)	0.6574	0.5827	0.5282	0.5, 97, 0.06	0.7831	0.3149	0.2932
	Bati	5-4-1	0.7914	0.2786	0.2645	(5,1,0)	0.6340	0.5595	0.5295	0.8, 98, 0.05	0.7681	0.2958	0.2804
<i>12 Month lead time</i>													
Upper	Bantu	5-4-1	0.7122	0.3545	0.3422	(4,1,1)	0.5442	0.7470	0.6946	0.55, 88, 0.08	0.7259	0.3256	0.3179
	Liben	6-4-1	0.7294	0.3620	0.3467	(5,1,0)	0.5078	0.7414	0.6379	0.65, 96, 0.08	0.7219	0.3282	0.3074
	Ginchi	6-4-1	0.7164	0.3632	0.3474	(5,0,10)	0.5939	0.7377	0.6200	0.85, 99, 0.06	0.7083	0.3387	0.3188
	Sebeta	7-4-1	0.7231	0.3640	0.3493	(5,1,0)	0.5025	0.7330	0.5876	0.8, 88, 0.09	0.7296	0.3667	0.3594
	Ejersalele Ziquala	5-4-1	0.7017	0.3692	0.3508	(5,0,1)	0.5145	0.7288	0.5804	0.6, 91, 0.06	0.7139	0.3786	0.3683
Middle	Modjo	5-4-1	0.7129	0.3816	0.3710	(4,0,1)	0.4780	0.8637	0.7777	0.6, 91, 0.05	0.7233	0.3783	0.3684
	Wolenchiti	5-4-1	0.7064	0.3705	0.3657	(5,0,0)	0.4471	0.8762	0.8186	0.45, 100, 0.09	0.7235	0.3630	0.3529
	Gelemsso	6-4-1	0.7215	0.3534	0.3290	(4,1,1)	0.5659	0.8889	0.8300	0.6, 98, 0.08	0.7242	0.3768	0.3710
	Dire Dawa	6-4-1	0.7146	0.3412	0.3148	(5,1,0)	0.5216	0.9035	0.7709	0.65, 92, 0.05	0.7023	0.3687	0.3433
Lower	Dubti	5-4-1	0.7179	0.3716	0.3580	(5,0,0)	0.5158	0.8854	0.7244	0.5, 88, 0.09	0.7253	0.3398	0.3158
	Eliwuha	6-4-1	0.7189	0.3715	0.3471	(4,0,1)	0.4265	0.8643	0.7275	0.25, 100, 0.1	0.7018	0.3420	0.3261
	Bati	6-4-1	0.7136	0.3708	0.3600	(5,1,0)	0.4915	0.8312	0.7729	0.88, 95, 0.07	0.7061	0.3387	0.2916

**Table 7**

The best WA-ANN and WA-SVR models for 6 and 12 month forecasts of SPI 24. Column 3 is the ANN architecture detailing the number of nodes in the input, hidden and output layers respectively. In column 7 the parameters of the SVR models are given.

Basin	Station	WA-ANN	R <sup>2</sup>	RMSE	MAE	WA-SVR	R <sup>2</sup>	RMSE	MAE
<i>6 Month lead time</i>									
Upper	Bantu Liben	5-6-1	0.9665	0.1968	0.1803	0.8, 93, 0.08	0.8832	0.2461	0.2108
	Ginchi	5-5-1	0.9254	0.2850	0.2671	0.8, 94, 0.06	0.8740	0.2475	0.2148
	Sebeta	5-5-1	0.8864	0.1821	0.1723	0.9, 93, 0.08	0.8742	0.2581	0.2331
	Ejersalele	7-5-1	0.8791	0.1778	0.1632	0.8, 93, 0.08	0.8683	0.2287	0.2023
	Ziquala	6-5-1	0.8978	0.2546	0.2395	0.5, 99, 0.05	0.8869	0.2192	0.2133
Middle	Modjo	5-4-1	0.9166	0.2561	0.2481	0.7, 92, 0.06	0.8938	0.2407	0.2281
	Wolenchiti	7-4-1	0.9014	0.2115	0.2049	0.6, 94, 0.08	0.8892	0.2218	0.2106
	Gelemsso	6-5-1	0.9407	0.2258	0.2054	0.45, 96, 0.01	0.8935	0.2773	0.2694
	Dire Dawa	7-6-1	0.9215	0.2335	0.2149	0.4, 89, 0.03	0.8865	0.2487	0.2124
Lower	Dubti	5-4-1	0.8953	0.2618	0.2531	0.4, 94, 0.07	0.8938	0.3094	0.2870
	Eliwuha	5-4-1	0.9122	0.2190	0.2015	0.5, 97, 0.06	0.9024	0.2100	0.1993
	Bati	6-4-1	0.9450	0.2236	0.2159	0.8, 98, 0.05	0.8953	0.2386	0.2114
<i>12 Month lead time</i>									
Upper	Bantu Liben	5-6-1	0.8372	0.3342	0.3025	0.55, 88, 0.08	0.8171	0.3531	0.3388
	Ginchi	6-5-1	0.8637	0.3373	0.3191	0.65, 96, 0.08	0.8284	0.2743	0.2414
	Sebeta	5-5-1	0.8331	0.3325	0.3282	0.85, 99, 0.06	0.8398	0.3548	0.3302
	Ejersalele	8-6-1	0.8277	0.3372	0.3362	0.8, 88, 0.09	0.8420	0.3353	0.3209
	Ziquala	7-5-1	0.8316	0.3245	0.3083	0.6, 91, 0.06	0.8554	0.3560	0.3422
Middle	Modjo	6-4-1	0.8419	0.3061	0.2931	0.6, 91, 0.05	0.8013	0.3591	0.3396
	Wolenchiti	8-5-1	0.8134	0.2948	0.2844	0.45, 100, 0.09	0.8400	0.3159	0.2760
	Gelemsso	7-6-1	0.8178	0.2879	0.2753	0.6, 98, 0.08	0.8223	0.2777	0.2697
	Dire Dawa	8-6-1	0.8659	0.2853	0.2762	0.65, 92, 0.05	0.8548	0.3410	0.3163
Lower	Dubti	5-5-1	0.8303	0.2827	0.2749	0.5, 88, 0.09	0.8472	0.3051	0.2827
	Eliwuha	6-4-1	0.8462	0.2874	0.2697	0.25, 100, 0.1	0.8494	0.3184	0.3073
	Bati	8-4-1	0.8086	0.2819	0.2620	0.88, 95, 0.07	0.8538	0.3019	0.2794

ANN models. For forecasts of SPI 12 (6 months lead time) the results are very similar across all stations. Fig. 6 shows the forecast results of the Eliwuha station for SPI 12. The figure indicates that the WA-ANN forecasts closely mirror the observed results and do not underestimate extreme events of precipitation or drought. The model does not accurately depict the inception of the drought period at the end of 1970. While the observed SPI series indicates a

slight drought, the forecasts indicate a moderate period of precipitation.

Fig. 7 is a scatter plot of the observed and predicted SPI values from the WA-ANN model. The figure shows that the points are closer to the trend line and there are no points of significant overestimation or underestimation. Fig. 8 illustrates the effects of the approximation series on the SPI time series for the Eliwuha station.

**Table 8**  
Results for the calibration/training data set of the Eliwuha station (SPI 12, 6 months lead time).

Model	R <sup>2</sup>	RMSE	MAE
ARIMA	0.5837	0.8903	0.8773
ANN	0.7241	0.3348	0.3180
SVR	0.7528	0.3062	0.2381
WA-ANN	0.9335	0.2140	0.1906
WA-SVR	0.8631	0.2481	0.2301

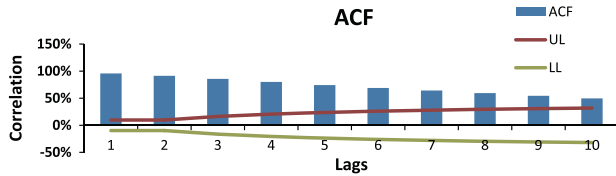


Fig. 3a. Autocorrelation of the SPI 12 time series for the Eliwuha station.

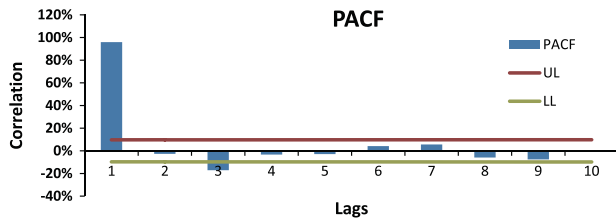


Fig. 3b. Partial autocorrelation of the SPI 12 time series for the Eliwuha station.

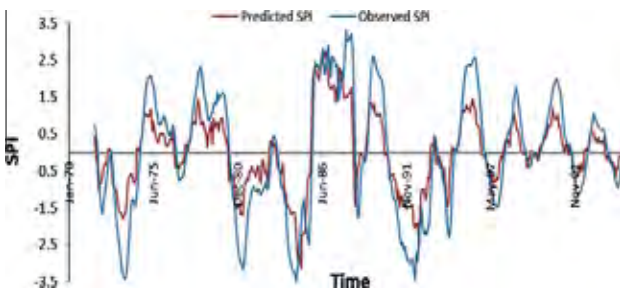


Fig. 4. Observed vs. predicted SPI 12 at Eliwuha station for the ANN model.

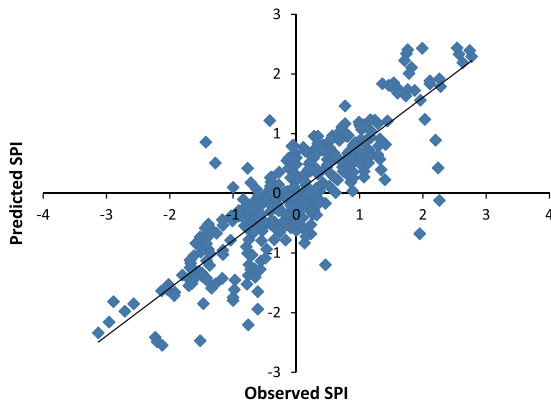


Fig. 5. Scatter plot for SPI values at Eliwuha station for the ANN model.

This study found that the approximation series after wavelet decomposition is useful to use in forecasting SPI time series. For the model in Fig. 8, the approximation series is from wavelet

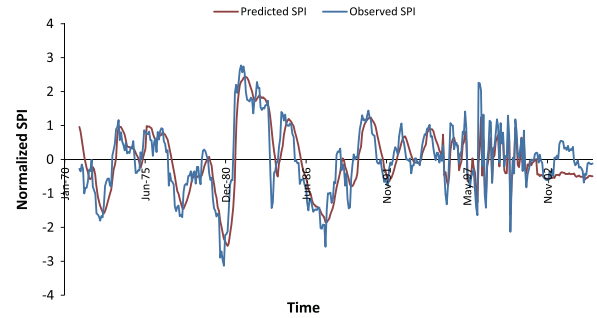


Fig. 6. Observed vs. predicted SPI at the Eliwuha station for the WA-ANN model.

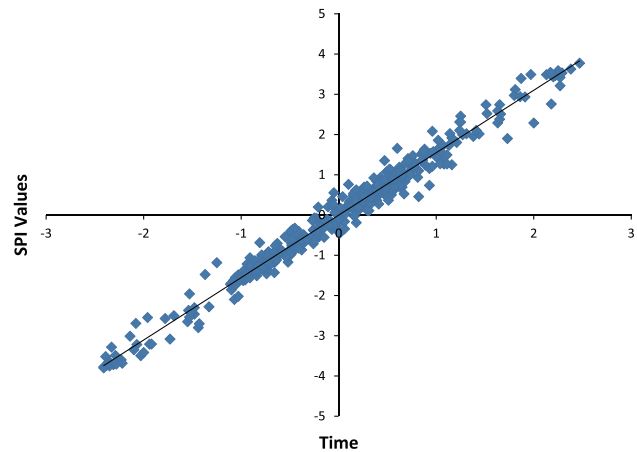


Fig. 7. Scatter plot from WA-ANN forecasts of SPI 12 at the Eliwuha station.

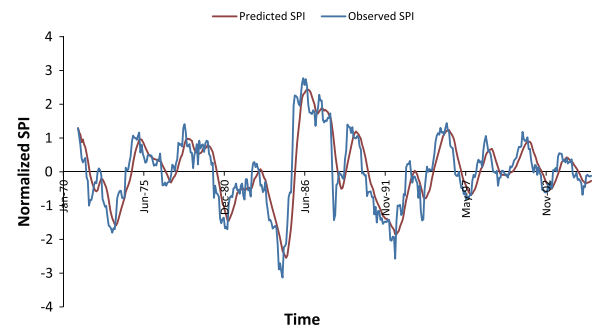


Fig. 8. Time series showing the observed SPI values against the approximation series after wavelet decomposition for the Eliwuha station.

decomposition at level three. The approximation series closely mirrors the periods of precipitation and drought exhibited by the SPI series. This approximation time series was subsequently used as an input in WA-ANN or WA-SVR models. As seen in Tables 3–6, the forecast results for WA-ANN and WA-SVR models are improved compared to models without any wavelet pre-processing.

5.4. SVR models

Fig. 9 illustrates the forecast results of the best proposed SVR model at the Eliwuha station. The SVR model had trouble predicting the inception of drought at the end of 1970. While the model does accurately predict events of drought there seems to be a slight time shift error. The predicted values are slightly lagged compared

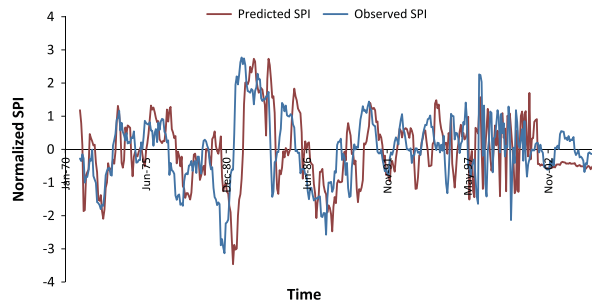


Fig. 9. Observed vs. predicted SPI 12 from a SVR model at the Eliwuha Station.

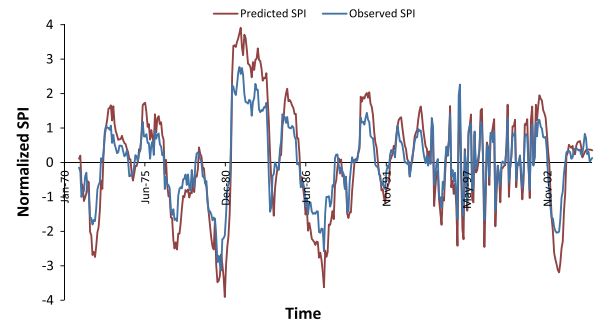


Fig. 11. Observed vs. predicted SPI 12 from a WA-SVR model at the Eliwuha Station.

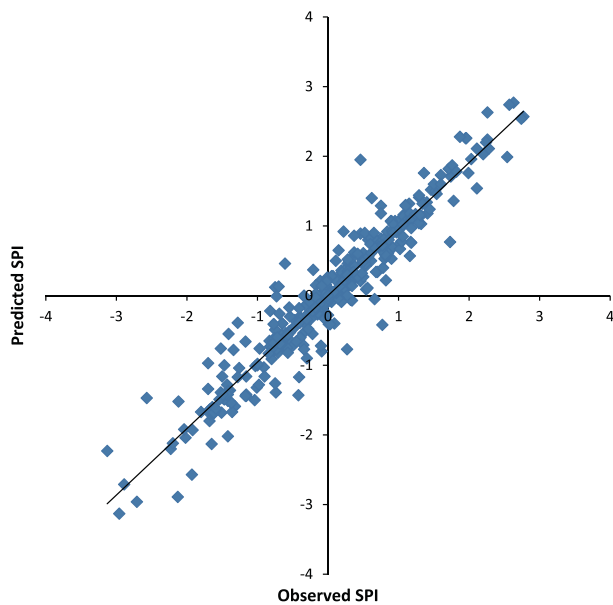


Fig. 10. Scatter plot of SPI 12 values for a SVR model at the Eliwuha Station.

to the observed values, which may make applications of SVR models for drought forecasting problematic. The SVR model also had a good correlation between the observed and predicted values as exhibited by the  $R^2$  value of 0.7382. The scatter plot in Fig. 10 illustrates the correlation between the observed and predicted SPI values. The observations in the scatter plot are symmetrical around the trend line.

### 5.5. WA-SVR models

The use of wavelet analysis improved the forecasting ability of SVR models with respect to SPI 12 and SPI 24. As shown in Tables 4 and 6, the use of wavelet analysis improves the performance measures across all the stations. The  $R^2$  values for the WA-SVR models are between 0.8144 and 0.8968; these values are all greater than any SVR model for SPI 12, indicating a greater level of correlation between observed and predicted values. For forecasts of 12 months lead time for both SPI values refer to Tables 3–6. As with other models an increase in lead time results in a deterioration of the forecast accuracy.

The forecast results from WA-SVR models are more accurate than SVR models, according to the performance measures. The inputs for the WA-SVR models were generated from the approximation series after wavelet decomposition of the SPI time series. The fact that the use of just the approximation series improved the results significantly indicates that the approximation series

adequately de-noises the data and avoids any discontinuities within the SPI time series.

Fig. 11 indicates that the WA-SVR model at the Eliwuha station is good at predicting the inception of a drought as shown in the figure around the end of 1970. Fig. 11 also indicates that the WA-SVR model tends to overestimate periods of extreme precipitation or drought. The overestimation is not a mis-categorization of the event. SPI values of  $-2$  and  $-3$  are both representative of extreme events. In general, the predicted values mirror the trends observed with the original time series. For the Eliwuha station, the WA-SVR model is effective at predicting the end of the SPI time series. The results from Figs. 6 and 9 show the models predicting slight to moderate drought while the observed time series indicates slight precipitation. The WA-SVR model was better at predicting the events at the end of the time series. The scatter plot in Fig. 12 shows the good level of correlation between the observed and predicted SPI values at the Eliwuha station.

## 6. Discussion

This study has shown that data driven models can be an effective means of forecasting drought at forecast lead times of 6 and 12 months in the Awash River Basin. The results indicate that machine learning techniques (ANNs and SVR) are more effective than a traditional stochastic model such as an ARIMA model in forecasting SPI 12 and SPI 24 at the aforementioned lead times. This is likely due to the fact that ANN and SVR models are effective in modeling non-linear components of time series data. Furthermore, the use of wavelet analysis as a pre-processing tool improved the forecast results for both ANN and SVR models. As might be expected, the results also indicate that as the forecast lead time is increased the correlation between observed and predicted values, as measured by  $R^2$ , decreases considerably. While the RMSE and the MAE increase with increasing forecast lead time, their increase is not as pronounced. An increase in forecast lead time from 6 to 12 months did not result in poor results, especially when wavelet analysis was used, which highlights the effectiveness of this pre-processing method for ANN and SVR models in predicting the SPI. The input structure of the models does not change with the increase of forecast lead time, which makes the models convenient for operational purposes.

The results from all the data driven models generally show that SPI 24 forecasts were more accurate than SPI 12 forecasts. Both SPI 12 and SPI 24 are long-term SPI and each new month has less impact on the period of sum precipitation (McKee et al., 1993) compared to short-term precipitation. As a result, monthly variation in precipitation has a smaller impact for both these SPI compared to short-term SPI. However, as SPI 24 is a longer term SPI its sensitivity to changes in precipitation within the time series is less than that of SPI 12. This lack of sensitivity may explain why the models are slightly more effective at generalizing SPI 24 better than SPI 12.



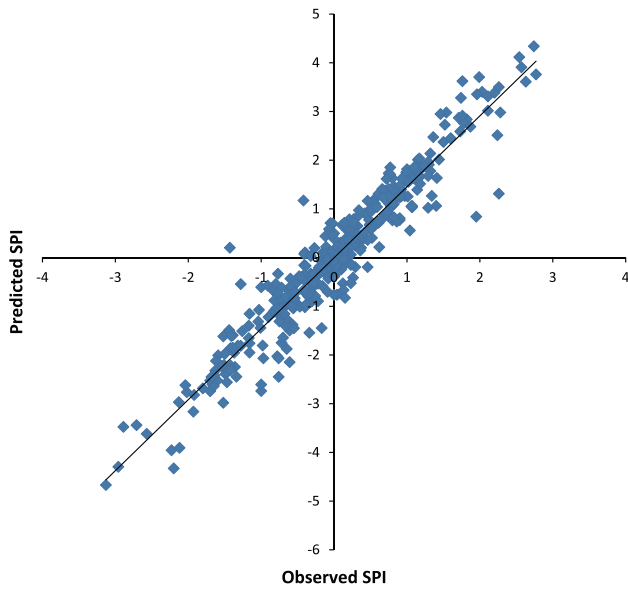


Fig. 12. Scatter plot from WA-SVR forecasts of SPI 12 at the Eliwuha station.

The forecast accuracy of the proposed models does not differ significantly between each of the basins. The lack of a significant difference in terms of the forecast accuracy indicates that the different conditions within the three sub-basins do not appreciably affect the forecast of the SPI. In general, the performances of SVR and ANN models were comparable. The use of wavelets improved the results of both machine learning techniques, with the forecast measures indicating that WA-ANN models slightly outperformed WA-SVR models. Theoretically, SVR models should perform better than ANN models because they adhere to the structural risk minimization principle instead of the empirical risk minimization principle. They should, in theory, not be as susceptible to local minima or maxima. However, there have been studies that have shown that the performance of SVR and ANN models are comparable. Lima et al. (2013) found that SVR models were more effective at precipitation forecasts when MAE was the performance measure and ANN models were more effective when MSE was the performance measure. A study by Shin et al. (2005) and Chevalier et al. (2011) found that the application of ANN models in time series forecasting was comparable to those of SVR models especially as the size of the training set was increased. The study by Chevalier et al. (2011) also found that SVR models were superior in the training phase, while ANN models were superior in the evaluation phase. Witten et al. (2011) found that ANN models are comparable to SVR models because they can learn to ignore irrelevant attributes. Witten et al. (2011) also state that there is no universally superior learning method.

From the figures illustrating all the forecasts using all of the data driven models, it is apparent that there is not much time step error in the forecasts of SPI 12. However, the SVR model at the Eliwuha station for SPI 12 did show a lag between observed and predicted events. This time shift error is unique to the SVR models in this study. This time shift error associated with the SVR models is indicated by the delayed drought forecasts. Time shift error is caused by the autoregressive components related to the selection of input variables (Abrahart et al., 2007). The use of past data in the forecasts of SPI values at long lead times introduces time shift error. Long lead time forecasts that do not possess time shift error usually result in noisier forecasts (Chua and Wong, 2011). In an operational setting, the presence of a time shift error would likely compromise the ability of planners to implement an effective drought warning system. Forecasts of SPI 12 at the Eliwuha station

using the SVR model showed that the inception of the drought (according to the observed time series) was not accurately predicted. However, the WA-SVR model at the same station did adequately predict the inception of drought but was relatively noisier than the SVR model.

## 7. Conclusion

This study investigated the ability of data driven models to forecast drought. This study also proposed and evaluated, for the first time, the use of the WA-SVR method for long-term drought forecasting.

Overall, coupled wavelet-neural network (WA-ANN) models were found to provide better results than the other model types used for forecasts of SPI 12 and SPI 24 in the Awash River Basin, especially for SPI 24. WA-ANN models showed a higher coefficient of determination between observed and predicted SPI, as well as lower RMSE and MAE values compared to simple ANNs, ARIMA, SVR and WA-SVR models. Wavelet coupled models also consistently showed lower values of RMSE and MAE compared to the other data driven models. The coupled models provide more accurate results because pre-processing the original SPI time series with wavelet decompositions improves the forecast results over time series that do not use wavelet decompositions. Wavelet analysis de-noises the SPI time series and subsequently allows the ANN and SVR model to model the main signal without the noise. Wavelet analysis also reduces the sensitivity to changes in monthly precipitation within the SPI time series. In the case of SVR models, the use of wavelets also reduces the lag between forecasts and the observed SPI values.

This study focused on long-term drought forecasts of SPI 12 and SPI 24 in the Awash River Basin. Further studies need to be done to determine which of these data driven models is suitable for forecasting long-term SPI values in other locations with different climates and different physical characteristics. Considering the fact that the Middle and Lower Awash sub-basins have a very similar climate, studies of areas with different climates should be conducted to determine whether there is a significant link between forecast accuracy and climate. This study found that the different characteristics and climatology of the sub-basins had no discernible effect on forecast accuracy. Future studies should also attempt to couple data driven drought forecasting models with uncertainty analysis, such as bootstrapping or boosting ensembles.

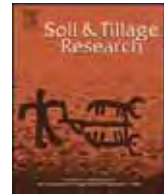
## Acknowledgements

This research was funded by an NSERC Discovery Grant and a CFI Grant held by Jan Adamowski. The data was obtained from National Meteorological Agency of Ethiopia. Their help is greatly appreciated.

## References

- Abrahart, R.J., Heppenstall, A.J., See, L.M., 2007. Timing error correction procedure applied to neural network rainfall/runoff modelling. *Journal of Hydrologic Sciences* 52 (3), 414–432.
- Adamowski, J., 2008. Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. *Journal of Hydrology* 353, 247–266.
- Adamowski, J., Chan, H.F., 2011. A wavelet neural network conjunction model for groundwater level forecasting. *Journal of Hydrology* 407, 28–40.
- Adamowski, J., Sun, K., 2010. Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology* 390, 85–91.
- Adamowski, J., Chan, E., Prasher, S., Ozga-Zielinski, B., Sliusareve, A., 2012. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resources Research* 48, W01528. <http://dx.doi.org/10.1029/2010WR009945>.

- Bacanli, U.G., Firat, M., Dikbas, F., 2008. Adaptive neuro-fuzzy inference system for drought forecasting. *Stochastic Environmental Research and Risk Assessment* 23 (8), 1143–1154.
- Barros, A., Bowden, G., 2008. Toward long-lead operational forecasts of drought: an experimental study in the Murray-Darling River Basin. *Journal of Hydrology* 357 (3–4), 349–367.
- Belayneh, A., Adamowski, J., 2012. Standard Precipitation Index drought forecasting using neural networks, wavelet neural networks, and support vector regression. *Applied Computational Intelligence and Soft Computing*. Article ID 794061, 13 pages.
- Beven, K., 2006. A manifesto for the equifinality thesis. *Journal of Hydrology* 320, 18–36.
- Bordi, I., Sutera, A., 2007. Drought monitoring and forecasting at large scale. In: Rossi, G. et al. (Eds.), *Methods and Tools for Drought Analysis and Management*. Springer.
- Box, G.E.P., Jenkins, G.M., 1976. *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco.
- Cacciamani, C., Morgillo, A., Marchesi, S., Pavan, V., 2007. Monitoring and forecasting drought on a regional scale: Emilia-Romagna region. *Water Science and Technology Library* 62 (1), 29–48.
- Cancelliere, A., Di Mauro, G., Bonaccorso, B., Rossi, G., 2007a. Stochastic forecasting of drought indices. In: Rossi, G. et al. (Eds.), *Methods and Tools for Drought Analysis and Management*. Springer.
- Cancelliere, A., Di Mauro, G., Bonaccorso, B., Rossi, G., 2007b. Drought forecasting using the standard precipitation index. *Water Resource Management* 21, 801–819.
- Cannas, B., Fanni, A., Sias, G., Tronci, S., Zedda, M.K., 2006. River flow forecasting using neural networks and wavelet analysis. *Proceedings of the European Geosciences Union*.
- Chevalier, R.F., Hoogenboom, G., McClendon, R.W., Paz, J.A., 2011. Support vector regression with reduced training sets for air temperature prediction: a comparison with artificial neural networks. *Neural Computing and Applications* 20, 151–159.
- Chua, I.H.C., Wong, T.S.W., 2011. Runoff forecasting for an asphalt plane by artificial neural networks and comparisons with kinematic wave and autoregressive moving average models. *Journal of Hydrology* 397, 191–201.
- Cimen, M., 2008. Estimation of daily suspended sediments using support vector machines. *Hydrological Sciences Journal* 53 (3), 656–666.
- Cutore, P., Di Mauro, G., Cancelliere, A., 2009. Forecasting palmer index using neural networks and climatic indexes. *Journal of Hydrologic Engineering* 14, 588–595.
- Desalegn, C., Babel, M.S., Das Gupta, A., Seleshi, B.A., Merrey, D., 2006. Farmers' perception about water management under drought conditions in the Awash River Basin, Ethiopia. *International Journal of Water Resources Development* 22 (4), 589–602.
- Edossa, D.C., Babel, M.S., Gupta, A.D., 2010. Drought analysis on the Awash River Basin, Ethiopia. *Water Resource Management* 24, 1441–1460.
- Gao, J.B., Gunn, S.R., Harris, J., Brown, M., 2001. A probabilistic framework for SVM regression and error bar estimation. *Machine Learning* 46, 71–89.
- Han, P., Wang, P., Zhang, S., Zhu, D., 2010. Drought forecasting with vegetation temperature condition index. *Wuhan Daxue Xuebao (Xinxi Kexue Ban)/Geomatics and Information Science of Wuhan University* 35 (10), 1202–1206+1259.
- Husak, G.J., Funk, C.C., Michaelsen, J., Magadzire, T., Goldsberry, K.P., 2013. Developing seasonal rainfall scenarios for food security early warning. *Theoretical and Applied Climatology*. <http://dx.doi.org/10.1007/s00704-013-0838-8>.
- Karamouz, M., Rasouli, K., Nazil, S., 2009. Development of a hybrid index for drought prediction: case study. *Journal of Hydrologic Engineering* 14, 617–627.
- Keshavarz, M., Karami, E., Vanclay, F., 2013. The social experience of drought in rural Iran. *Journal of Land Use Policy* 30, 120–129.
- Khan, M.S., Coulibaly, P., 2006. Application of support vector machine in lake water level prediction. *Journal of Hydrologic Engineering* 11 (3), 199–205 (American Society of Civil Engineering).
- Kim, T., Valdes, J.B., 2003. Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. *Journal of Hydrologic Engineering* 8, 319–328.
- Kisi, O., Cimen, M., 2009. Evapotranspiration modelling using support vector machines. *Hydrological Science Journal* 54 (5), 918–928.
- Kisi, O., Cimen, M., 2011. A wavelet-support vector machine conjunction model for monthly streamflow forecasting. *Journal of Hydrology* 399, 132–140.
- Labat, D., Ababou, R., Mangin, A., 1999. Wavelet analysis in Karstic hydrology 2nd Part: rainfall-runoff cross-wavelet analysis. *Comptes Rendus de l'Academie des Sciences Series IIA Earth and Planetary Science* 329, 881–887.
- Lane, S.N., 2007. Assessment of rainfall-runoff models based upon wavelet analysis. *Hydrological Processes* 21, 586–607.
- Lima, A.R., Cannon, A.J., Hsieh, W.W., 2013. Nonlinear regression in environmental sciences by support vector machines combined with evolutionary strategy. *Computers and Geosciences* 50, 136–144.
- Linsley, R.K., Kohler, M.A., Paulhus, J.L.H., 1988. *Hydrology for Engineers*, International Edition. McGraw-Hill, Singapore.
- Maheswaran, R., Khosa, R., 2012a. Comparative study of different wavelets for hydrologic forecasting. *Computers and Geosciences* 46, 284–295.
- Maheswaran, R., Khosa, R., 2012b. Wavelet-volterra coupled model for monthly streamflow forecasting. *Journal of Hydrology* 450, 320–335.
- Mallat, S.G., 1998. *A Wavelet Tour of Signal Processing*. Academic, San Diego, p. 577.
- Marj, A.F., Meijerink, A.M., 2011. Agricultural drought forecasting using satellite images, climate indices and artificial neural network. *International Journal of Remote Sensing* 32 (24), 9707–9719.
- McKee, T.B., Doesken, N.J., Kleist, J., 1993. The Relationship of Drought Frequency and Duration to Time Scales, Paper Presented at 8th Conference on Applied Climatology. American Meteorological Society, Anaheim, CA.
- Mishra, A.K., Desai, V.R., 2005. Drought forecasting using stochastic models. *Stochastic Environmental Research and Risk Assessment* 19 (5), 326–339.
- Mishra, A.K., Desai, V.R., 2006. Drought forecasting using feed-forward recursive neural network. *Ecological Modelling* 198 (1–2), 127–138.
- Mishra, S.S., Nagarajan, R., 2012. Forecasting drought in Tel River Basin using feed-forward recursive neural network. *International Conference on Environmental, Biomedical and Biotechnology*.
- Mishra, A.K., Singh, V.P., 2012. Simulating hydrological drought properties at different spatial units in the United States based on wavelet-bayesian approach. *Earth Interactions* 16 (17), 1–23.
- Mishra, A.K., Desai, V.R., Singh, V.P., 2007. Drought forecasting using a hybrid stochastic and neural network model. *Journal of Hydrologic Engineering* 12, 626–638.
- Morid, S., Smakhtin, V., Moghaddasi, M., 2006. Comparison of seven meteorological indices for drought monitoring in Iran. *International Journal of Climatology* 26 (7), 971–985.
- Morid, S., Smakhtin, V., Bagherzadeh, K., 2007. Drought forecasting using artificial neural networks and time series of drought indices. *International Journal of Climatology* 27 (15), 2103–2111.
- Murtagh, F., Starc, J.L., Renaud, O., 2003. On neuro-wavelet modeling. *Decision Support Systems* 37, 475–484.
- Nason, G.P., Von Sachs, R., 1999. Wavelets in time-series analysis. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 357 (1760), 2511–2526.
- Ntale, H.K., Gan, T.Y., 2003. Drought indices and their application to East Africa. *International Journal of Climatology* 23 (11), 1335–1357.
- Ozger, M., Mishra, A.K., Singh, V.P., 2012. Long lead time drought forecasting using a wavelet and fuzzy logic combination model: a case study in Texas. *Journal of Hydrometeorology* 13 (1), 284–297.
- Parrella, F., 2007. *Online Support Vector Regression*. Master Thesis, University of Genoa.
- Partal, T., Kisi, O., 2007. Wavelet and neuro-fuzzy conjunction model for precipitation forecasting. *Journal of Hydrology* 342 (1–2), 199–212.
- Renaud, O., Starc, J., Murtagh, F., 2002. Wavelet-based Forecasting of Short and Long Memory Time Series. Department of Economics, University of Geneva.
- Shin, K., Lee, T.S., Kim, H., 2005. An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications* 28, 127–135.
- Tiwari, M.K., Chatterjee, C., 2010. Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach. *Journal of Hydrology* 394, 458–470.
- Tsakiris, G., Vangelis, H., 2004. Towards a drought watch system based on spatial SPI. *Water Resources Management* 18 (1), 1–12.
- Vapnik, V., 1995. *The Nature of Statistical Learning Theory*. Springer Verlag, New York, USA.
- Wanas, N., Auda, G., Kamel, M.S., Karray, F., 1998. On the optimal number of hidden nodes in a neural network. *Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering*, 918–921.
- Wilhite, D.A. (Ed.), 1993. *Drought Assessment, Management, and Planning: Theory and Case Studies*. Natural Resource Management and Policy Series. Kluwer Publishers, Boston, MA.
- Witten, I.H., Frank, E., Hall, M.A., 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufman, Burlington, MA.



# Soil failure patterns and draft as influenced by consistency limits: An evaluation of the remolded soil cutting test



A.A. Tagar<sup>a,b</sup>, Changying Ji<sup>a,\*</sup>, Qishuo Ding<sup>a</sup>, Jan Adamowski<sup>c</sup>, F.A. Chandio<sup>a,b</sup>, I.A. Mari<sup>a</sup>

<sup>a</sup> College of Engineering, Nanjing Agricultural University, Nanjing 210031, PR China

<sup>b</sup> Faculty of Agricultural Engineering, Sindh Agriculture University, Tandojam 70060, Pakistan

<sup>c</sup> Department of Bioresource Engineering, McGill University, Ste Anne de Bellevue, Quebec H9X3V9, Canada

## ARTICLE INFO

### Article history:

Received 2 May 2013

Received in revised form 3 December 2013

Accepted 5 December 2013

### Keywords:

Sticky limit

Sticky point

Flat triangular shaped tool

Dry land soil

Paddy soil

Laboratory soil bin

## ABSTRACT

Soil failure patterns play an important role in obtaining a better understanding of the mechanical behavior of soils. Despite the large number of studies over the past few decades, a better understanding of soil failure patterns and its relation to soil and tool parameters for particular soils such as dry land and paddy has not been developed. This study investigated soil failure patterns and related draft at sticky, plastic and liquid consistency limits and the sticky point of dry land and paddy soils. A soil cutting test rig was developed to perform soil cutting at three consistency limits (sticky limit, plastic limit and liquid limit) and the sticky point of soil, three rake angles (15°, 30° and 45°), and three operating depths (30 mm, 50 mm and 70 mm). A flat triangular shaped tool operating at a constant speed of 10 mm s<sup>-1</sup> was used in all experiments. Soil failure patterns were observed and recorded using a digital camera, and draft per unit displacement was measured by load cells attached to the soil bin. A direct relationship between soil failure patterns or draft and the consistency limits of soil was found. Brittle failure was observed at the sticky limit, chip forming failure was observed at 15° rake angle and 30 mm depth, and bending failure with little strains of elements at 30° and 45° rake angles and 50 mm and 70 mm depths at plastic limit, while flow failure was linked to the liquid limit of the soil. At the sticky point, flow failure was observed at an operating depth of 30 mm and 15° rake angle, while flow with considerable bending and no strains of elements occurred at 50 mm and 70 mm operating depths and 30° and 45° rake angles. However, bending was more prominent at 70 mm depth and 45° rake angle. The draft at the sticky limit, plastic limit and sticky point was cyclic in nature, whereas at the liquid limit it was comparatively diverse and fading. The highest draft was found at the plastic limit, and the lowest at the liquid limit. Since the soil failure patterns may change with moisture content, soil type and particle size distribution within the same textural class, consistency limits can provide clearer and more accurate definitions of soil failure patterns than moisture content levels alone.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

The proper design and selection of soil-engaging tools to achieve desired soil tillage depends largely on the mechanical behavior of the soils (Rajaram and Erbach, 1998). Soil failure patterns play an important role in obtaining a better understanding of the mechanical behavior of soils under varied soil and tool conditions. The variation in soil failure patterns can be attributed to the variations of mechanical behavior of the soil (Abo Al-kheer et al., 2011). Soil failure patterns can include collapse, fracturing (brittle), chip forming, and flow (Salokhe, 1986; Rajaram and

Gee-Clough, 1988; Sharma, 1990). These failure patterns may vary with soil and tool parameters (Elijah and Weber, 1971; Godwin and Spoor, 1977; Stafford, 1981; Makanga et al., 1996).

Over the past couple of decades, numerous studies have been conducted to evaluate the effects of soil and tool parameters on soil failure patterns such as moisture content (Makanga et al., 2010; Rajaram and Gee-Clough, 1988; Wang and Gee-Clough, 1993), rake angles (Aluko and Seig, 2000), aspect ratio (Makanga et al., 1996) and operating speed (Stafford, 1979; Karmakar, 2005). Despite this large number of studies, a better understanding of the relationships existing between soil failure patterns and soil and tool parameters has not been elucidated. This is particularly true for dry land and paddy soils. This is likely mainly attributable to the complexity of soil genesis, textures, unique weather conditions, and the cropping systems associated with each soil under study.

\* Corresponding author. Tel.: +86 25 58606571.

E-mail addresses: [chyji@njau.edu.cn](mailto:chyji@njau.edu.cn), [chyji@yahoo.com](mailto:chyji@yahoo.com) (C. Ji).

Although the importance of a better understanding of the true failure patterns of soils has been emphasized by a number of authors (e.g., Rajaram and Erbach, 1997, 1998), technical methods to quantify soil failure patterns are limited. Moreover, the numerical value of the moisture content does not show any direct relationship with changes in soil failure patterns in different soils (Jayasuriya and Salokhe, 2001). One potential indicator may relate to a soil's consistency limits. The consistency limits of soil or Atterberg limits (Atterberg, 1911) define the resistance of soils to deformation or rupture. The plastic limit refers to the minimum moisture content at which soil can be puddled (Lal and Shukla, 2004), the liquid limit is the minimum moisture content at which soil flows like a liquid (Dexter and Bird, 2001; SSSA, 2009), and the sticky limit is the moisture content at which soil has little or no stickiness/adhesion to a steel spatula (Lal and Shukla, 2004). These terms are commonly used for classifying cohesive soils for engineering purposes (McBride, 2008). To some extent, they may yield information on the mechanical behavior of soils (Keller and Dexter, 2012). Soil workability is clearly related to the consistency limits of soil (Mueller, 1985; Smedema, 1993; Müller and Schindler, 1998a; Dexter and Bird, 2001; Müller et al., 2003). Archer (1975) concluded that consistency limits can be used to classify soils for workability. Moreover, Yamamoto (1963) suggested that data on the consistency limits of soil can be used as a guide for tillage practices. The sticky point, defined as the moisture content at which soil has maximum stickiness/adhesion to a steel spatula (Braja, 2002), is also an important criteria of soil workability. According to Müller and Schindler (1998b) the stickiness of soil is the limiting factor for its workability.

Many studies have explored soil failure patterns (e.g., Makanga et al., 2010; Wang and Gee-Clough, 1993; Rajaram and Gee-Clough, 1988; Stafford, 1981; Elijah and Weber, 1971). All studies to date have been based on specific soil moisture content levels. Although a few moisture levels could be inferred to represent the agronomically important soil moisture contents at the permanent wilting point, field capacity and saturation (e.g., Makanga et al., 2010), the rationale for choosing specific moisture content values was not specified in these studies. However, some studies provided the plastic limit and liquid limit of their soils, but unfortunately the experimental soil moisture levels that were used did not correspond to moisture content levels at plastic limit or liquid limit. In contrast, our study explores for the first time the use of soil consistency limits to investigate soil failure patterns during soil cutting. In order to provide precise definitions of soil failure patterns, this study was designed to investigate: (i) the relationship between soil failure patterns and draft with consistency limits and (ii) failure patterns and draft at the sticky point, in both dry land and paddy soils.

## 2. Materials and methods

The experiments in this study were carried out in an indoor soil bin test rig developed at the Department of Agricultural

Mechanization, College of Engineering, at Nanjing Agricultural University (NJAU) in China. The soils used in the experiment were dry land soil and paddy soil. These are yellow-brown soils according to the Chinese Soil Taxonomy and Halpudalf in the US classification scheme. China's yellow-brown soils, shown to have arisen through an independent soil genesis mechanism, are distributed across a wide swath of agro-ecological regions (27°33° N lat.), and have important implications for agricultural production. Consequently, an investigation of the general mechanical behaviors of these types of soil during cutting is warranted, particularly with respect to clearly defining its failure patterns. The paddy soil was used for a rice-wheat rotation on the university's Jiangpu Experimental Farm, while the farm's dry land soil was used to cultivate vegetables such as potatoes (*Solanum tuberosum* L., tomatoes (*Solanum lycopersicum* L.), eggplants (*Solanum melongena* L.) and chilies (*Capsicum* sp.).

### 2.1. Soil preparation

The experimental soil was air dried for two to three weeks, ground, and sieved through a 4 mm sieve. Composite soil samples were taken from the sieved soil to determine the moisture content in the soil, and then on the basis of existing moisture content a calculated amount of water was added to the soil (Eq. (1)) as a fine spray so as to attain the desired moisture content at the required consistency limit and the sticky point of the soil.

$$V_a = V_{req} - V_{ex} = (SMC_{req} \times W_s) - (SMC_{ex} \times W_s) \quad (1)$$

where,  $V_{req}$  is the volume of water which needs to be added to the soil in order to achieve the desired soil moisture content (ml),  $V_{ex}$  is the existing water present initially in the dry sieved soil (ml),  $SMC_{req}$  is the required moisture content ( $\text{g kg}^{-1}$ ),  $SMC_{ex}$  is the moisture content of the dry sieved soil ( $\text{g kg}^{-1}$ ), and  $W_s$  is the weight of soil (g).

It was then mixed well to obtain a homogenous soil specimen, covered with a polyethylene sheet to prevent evaporation, and left for 24 h to equilibrate to uniform moisture content. The soil was transferred to a metal-framed mold and compacted to the ideal bulk densities i.e., 1.22–1.4  $\text{mg m}^{-3}$  for sandy clay loam soil (USDA, 1999) using a manually operated compactor. If this was not done, soil compaction could have had effects on soil failure patterns. Soil molds at the sticky limit were compacted to 1.22–1.25  $\text{mg m}^{-3}$ . However, soil molds at the plastic limit, liquid limit and sticky point were compacted to 1.3–1.4  $\text{mg m}^{-3}$  (Table 1). Because of high moisture content, it was not possible to compact soil molds at the required density with the compaction device. This is consistent with Smith et al. (1997), who concluded that when the moisture content is so high that all the soil pores are filled with water, the soil becomes less compressible. These soil molds (300 mm × 100 mm × 100 mm) were then transferred to the soil cutting test rig.

**Table 1**

The physical and mechanical properties of dry land and paddy soils at different consistency limits and the sticky point.

Consistency limit	Dry land soil					Paddy soil				
	Moisture content ( $\text{g kg}^{-1}$ )	Dry bulk density ( $\text{mg m}^{-3}$ )	Cohesion (kPa)	Angle of internal friction (Deg.)	Cone index (kPa)	Moisture content ( $\text{g kg}^{-1}$ )	Dry bulk density ( $\text{mg m}^{-3}$ )	Cohesion (kPa)	Angle of internal friction (Deg.)	Cone index (kPa)
Sticky limit	15	1.22	0.0012	26	106	17.7	1.25	0.001	29	109
Plastic limit	22	1.32	0.0005	16	104	32	1.4	0.0004	18	106
Liquid limit	36	1.3	0	4	11	45	1.4	0	4	13
Sticky point	32	1.4	0.00005	8	18	44	1.3	0	4	15



**Table 2**

The physical and mechanical properties of dry land and paddy soils after different failure patterns.

Consistency limit	Type of failure pattern	Dry land soil				Paddy soil			
		Dry bulk density ( $\text{mg m}^{-3}$ )	Cohesion (kPa)	Angle of internal friction (Deg.)	Cone index (kPa)	Dry bulk density ( $\text{mg m}^{-3}$ )	Cohesion (kPa)	Angle of internal friction (Deg.)	Cone index (kPa)
Sticky limit	Brittle failure	1.06	0.0006	20.5	38.78	1.08	0.0004	26	39.33
Plastic limit	Chip forming	1.33	0.0004	15.5	36	1.32	0.0003	19	34.78
	Bending Failure	1.39	0.0006	16	42	1.37	0.0005	19	40
Liquid limit	Flow failure	1.3	0	5.5	4.44	1.31	0	4.5	8.22
Sticky point	Flow and bending	1.39	0	5.2	6	1.31	0	4.6	6

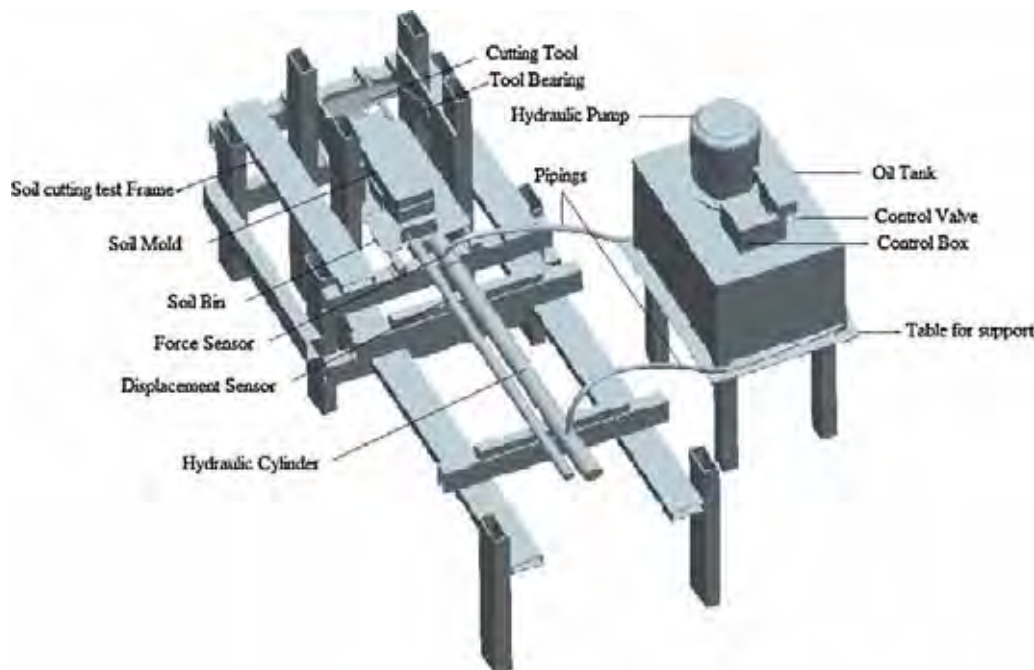
## 2.2. Soil physical and mechanical properties

Soil texture was determined by the Bouyoucos Hydrometer Method (Bouyoucos, 1927), soil moisture content and dry bulk density by the gravimetric method (Blake and Hartge, 1986), shear strength properties using a direct shear box apparatus (Fredlund and Vanapalli, 2002). The cone index (penetration resistance) was directly measured with a digital soil compaction meter (TJSD-750, Zhejiang Top Instrument Co. Ltd, China). The dry land soil was composed of 67% sand, 28% clay and 5% silt, and was classified as sandy clay loam, while the paddy soil was comprised of 50% sand, 24% clay, and 26% silt, and classified as sandy clay loam according to the USDA textural classification. The sticky limit of soil was measured as gravimetric moisture content at which soil fails to stick or adhere to a stainless steel spatula when the spatula blade is drawn across the face of the moist unknaded mass of soil, exerting a firm pressure against the soil (Baruah and Barthakur, 1997). The plastic limit was determined as the gravimetric moisture content at which a soil just begins to crumble as it is rolled into a thread of 3 mm diameter (Sowers, 1965). A 30° cone mounted on a shaft, with a total weight of about 80 g, was allowed to drop on a cup full of soil for 5 s. Soil moisture content corresponding to a penetration of 20 mm on the linear relationship between soil moisture content ( $x$ -axis) and the penetration ( $y$ -axis) was determined and considered as the cone penetrometer liquid limit

(Campbell, 2001). The sticky point of soil was measured at the point at which the soil had maximum stickiness/adhesion to steel spatula (Baver, 1956). The gravimetric soil moisture content at the respective consistency limits and sticky point differed between the two soils. The organic carbon content of dry land soil was  $8.4 \text{ g kg}^{-1}$ , and for paddy soil it was  $9.6 \text{ g kg}^{-1}$ . The physical and mechanical properties of dry land and paddy soils before the cutting test and after different failure patterns are shown in Tables 1 and 2.

## 2.3. Soil cutting test rig

A safe and easy to operate test rig was developed for the study. Its main features include a soil bin, tool bearing, cutting tool, hydraulic system, load cells, and a computer based data acquisition and control system. The rectangular steel tubings ( $90 \text{ mm} \times 40 \text{ mm} \times 5 \text{ mm}$ ), were cut into two sizes (i.e.,  $8 \text{ mm} \times 1000 \text{ mm}$  and  $16 \text{ mm} \times 600 \text{ mm}$ ), and were joined through bolts to construct a 2000 mm long and 600 mm wide test rig (Fig. 1). Two railings were provided on the test rig; a soil bin (500 mm long and 300 mm wide) with two supports ( $90 \text{ mm} \times 40 \text{ mm} \times 50 \text{ mm}$ ) on the right side and one at the front was mounted on the railings and fixed through steel screws. The height of the supports was kept smaller to avoid boundary frictions. Two steel sheets ( $200 \text{ mm} \times 150 \text{ mm} \times 10 \text{ mm}$ ) were fixed in the test rig at a distance of

**Fig. 1.** Soil cutting test rig used in the experiments.

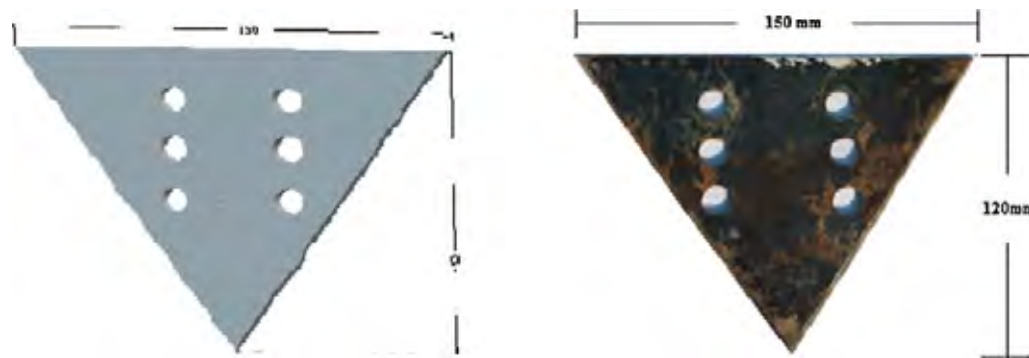


Fig. 2. Cutting tool used in the experiments.

100 mm on the left and right side of the soil bin, and the tool bearing was attached to them. The holes (15 mm diameter) were drilled into steel sheets at different angles so that their rake angles could be adjusted according to the conditions tested as done by Nalavade et al. (2010), Aluko and Seig (2000), and Rajaram and Erbach (1998). In order to move the soil bin to perform cutting operations, the soil bin was connected to the hydraulic system powered by an electric motor (2.2 kW) through a hydraulic cylinder. A two way solenoid valve was provided to move the soil bin (i) to perform cutting operations and (ii) to take it back to its original position. Two load cells connected to a computer were attached to the soil bin and a program was developed using LabVIEW (National Instruments, Austin, TX, USA) for data acquisition. The load cell designated as LSR-2A (2KN, Shanghai, Zhendan Sensor and Instrument Factory, China) was fixed in the hydraulic cylinder and attached to the soil bin to measure draft, whereas the load cell designated as W-DCD400 (1000 mm, Hefei, Gaochuang Sensor Co. Ltd. China) was placed near the hydraulic cylinder on the test rig and was attached to the soil bin to measure displacement. These load cells were calibrated prior to the experiments, similarly as described by Aluko and Seig (2000), and Makanga et al. (2010).

#### 2.4. Soil cutting test procedure

The soil cutting test was performed at three consistency limits (sticky limit, plastic limit and liquid limit) and at the sticky point of the soil, factorially combined with three rake angles (15°, 30° and 45°) and three operating depths (30 mm, 50 mm and 70 mm), as done by Wang (1991), Makanga et al. (1996), and Aluko and Seig (2000). A flat triangular shaped tool operating at a constant speed of 10 mm s<sup>-1</sup> was used in all experiments to allow for a clear observation and recording of soil failure patterns (Makanga et al., 1996, 2010). In order to adjust the cutting tool at different depths, three rows of holes (15 mm diameter), two in each row with 20 mm spacing, were drilled about 20 mm from the top at the center of the cutting tool (Fig. 2).

The soil molds were placed in the soil bin. As the cutting tool moved in the soil bin, the resulting soil failure patterns were observed and recorded using a digital camera (Canon Power Shot A4000 IS: 16 Mega pixels, Canon Inc., China). The recorded videos were then converted to snapshots. Simultaneously, the force-displacement signals were recorded through a data acquisition system (High Speed Advantech Bus Multifunction Data Acquisition Card USB-4711A with a resolution of 12 bits per channel sampling rate up to 150 ks/s) in an output file in the computer loaded with LabVIEW. This was then converted to MS-Excel spreadsheets to calculate draft along with the distance traveled per unit millimeter of 300 mm soil mold.

### 3. Results and discussion

#### 3.1. Soil failure patterns at different consistency limits and the sticky point

Soil failed in a brittle failure pattern in both dry land and paddy soils at all operating depths (30 mm, 50 mm and 70 mm) and rake angles (15°, 30° and 45°) at the sticky limit of soil (Fig. 3). However, in dry land soil the size of clod separated from the soil mold was 116 mm long, 49 mm wide and 100 mm thick; in the paddy soil the size of clod separated from the soil mold was 209 mm long, 62 mm wide and 100 mm thick. The above results showed a direct relationship between brittle failure at the sticky limit of both dry land and paddy soils. The close relation of the brittle failure pattern with the sticky limit demonstrates that, if brittle failure is desired (Dexter, 1975; Towner, 1987; Braunack et al., 1979; Snyder and Miller, 1989; Mullins et al., 1990; Okunlola and Payne, 1991; Aluko and Seig, 2000), soil should be ploughed at its sticky limit. Although this type of failure has been reported in a number of studies (Drees, 1956; Selig and Nelson, 1964; Elijah and Weber, 1971; Koolen, 1972, 1973; Rajaram and Gee-Clough, 1988; Rajaram and Erbach, 1998; Aluko and Seig, 2000; Aluko and Chandler, 2004), the soil conditions in those studies were not properly described.

Rajaram and Erbach (1996) pointed out that brittle failure or cracking of the soil in front of the tool occurs in moist soils, with moisture contents not exceeding the plastic limit. Similarly, Jayasuriya and Salokhe (2001) stated that brittle failure occurs in frictional-cohesive and cohesive soils when the moisture content is closer to (below) the plastic limit. However, the moist conditions closer to (below) are also difficult to recognize. According to Rajaram and Gee-Clough (1988), brittle failure occurs at 18.3% moisture content below plastic limit (PL, 22%). In contrast, Stafford (1981) obtained brittle failure pattern at 18% and 28% moisture content below and above the plastic limit (PL, 21%). Moreover, Makanga et al. (2010) found progressive shear type failure patterns at 21% moisture content below plastic limit (PL, 23%). Similarly, Harison (1982) observed progressive shear type failure at 18–20% moisture contents. In our study, brittle failure was found at the sticky limit of the soil.

A chip-forming failure pattern was found at 15° rake angle and 30 mm depth at the plastic limit in both dry land and paddy soils. The chip was 43 mm wide and 30 mm thick in dry land soil, while in paddy soil the chip was 62 mm wide and 30 mm thick. Rajaram (1987) and Rajaram and Gee-Clough (1988) observed chip formation failure at 28.6% moisture content above the plastic limit (PL, 22%) or between the plastic limit and saturation with vertical tines of varying width (30, 120 and 150 mm) at 100 mm depth. The bending failure with little strains of elements in the vertical direction was witnessed at 30° and 45° rake angles and

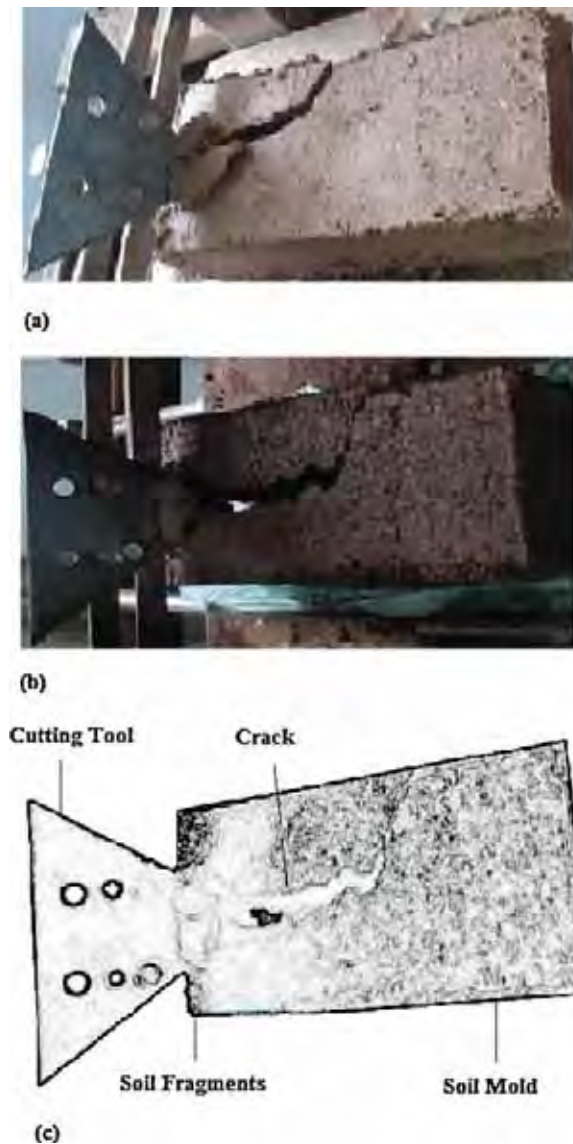


Fig. 3. Brittle failure at the sticky limit: (a) dry land soil, (b) paddy soil and (c) schematic sketch.

50 mm and 70 mm depths (Figs. 4 and 5). In dry land soil, the bending failure was 125 mm wide and 56 mm thick, while in paddy soil the bending failure was 106 mm wide and 53 mm thick. Elijah and Weber (1971) observed bending failure with inclined flat blades at 150 mm depth and a 45° rake angle above the plastic limit. At 70 mm depth and 45° rake angle, bending failure occurred in a reverse or backward direction. This indicated that at the plastic limit, chip forming failure may be observed at <50 mm depths and <30° rake angles, while bending failure with little strains of elements in the vertical direction occurred at ≥50 mm operating depths and ≥30° rake angles.

At the liquid limit, flow failure was observed in both soils at all operating depths (30 mm, 50 mm and 70 mm) and rake angles (15°, 30° and 45°) as shown in Fig. 6. Flow failure was 57 mm wide and 33 mm thick in dry land soil, while in paddy soil it was 43 mm wide and 32 mm thick. This type of failure pattern was also observed by Wang (1991) who conducted an experiment with a wide tine with varying rake angles (25–125°) in a Bangkok clay soil at 44% and 52% soil moisture content below the liquid limit (LL, 54%). He observed flow type soil failure at both soil moisture contents. Similarly, Rajaram and Gee-Clough (1988) observed

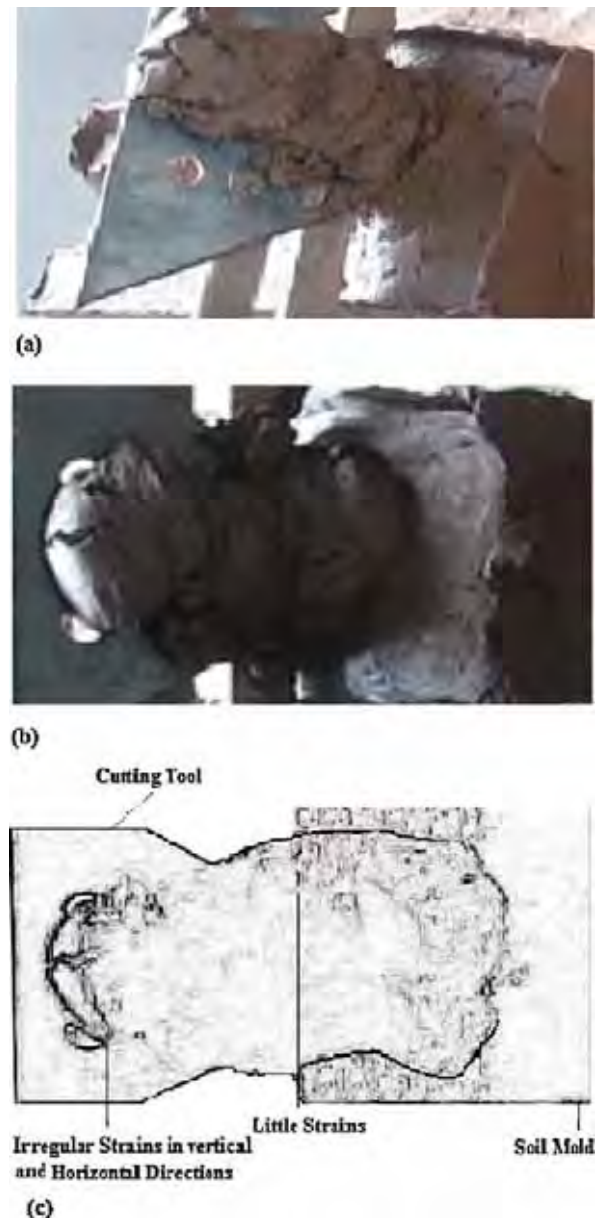


Fig. 4. Chip forming failure at the plastic limit: (a) dry land soil, (b) paddy soil and (c) schematic sketch.

flow-type soil failure at 42% below the liquid limit (LL, 46%). In contrast, Wang and Gee-Clough (1993), in studying a Bangkok clay soil (wet), found that at a soil moisture content as high as 44%, soil brittle failure was found at small and medium rake angles (<50°), as well as soil shear failure pattern. However, our study found that flow failure is clearly related to the liquid limit of the soil.

At the sticky point flow failure was observed at 30 mm depth and 15° rake angle, while flow with considerable bending and no strains of elements was experienced at 50 mm and 70 mm operating depths and 30° and 45° rake angles, as can be seen from Fig. 7. In dry land soil, the flow with considerable bending failure was 100.5 mm wide and 57 mm thick, while in paddy soil the flow with considerable bending failure was 108 mm wide and 55 mm thick. However, bending was more prominent at 70 mm depth and 45° rake angle, possibly because during cutting at 30 mm depth and 15° rake angle, the soil in contact with the cutting tool is less than compared to that of 50 mm and 70 mm operating depths and 30° and 45° rake angles. It is anticipated that





Fig. 5. Bending failure at the plastic limit: (a) dry land soil, (b) paddy soil and (c) schematic sketch.

if soil is ploughed at the sticky point at  $>70$  mm operating depths and  $>45^\circ$  rake angles, and at a speed of  $10 \text{ mm s}^{-1}$ , bending will be high enough to stop further soil cutting, because of the high stickiness and greater amount of soil that is in contact with the cutting tool.

Although flow failure (Stafford, 1981; Rajaram and Gee-Clough, 1988; Wang, 1991; Makanga et al., 1996, 2010) and bending failure patterns (Elijah and Weber, 1971) have been explored before in the literature, flow failure with considerable bending and no strains of elements at the sticky point are reported for the first time in the present study.

### 3.2. Draft at different consistency limits and the sticky point

The draft at the sticky limit (Fig. 8), plastic limit (Fig. 9) and sticky point (Fig. 11) were cyclic in nature. The cyclicity could be related to the observed repetitive and periodic nature of the soil failure patterns at different stages of tool travel (Makanga et al.,

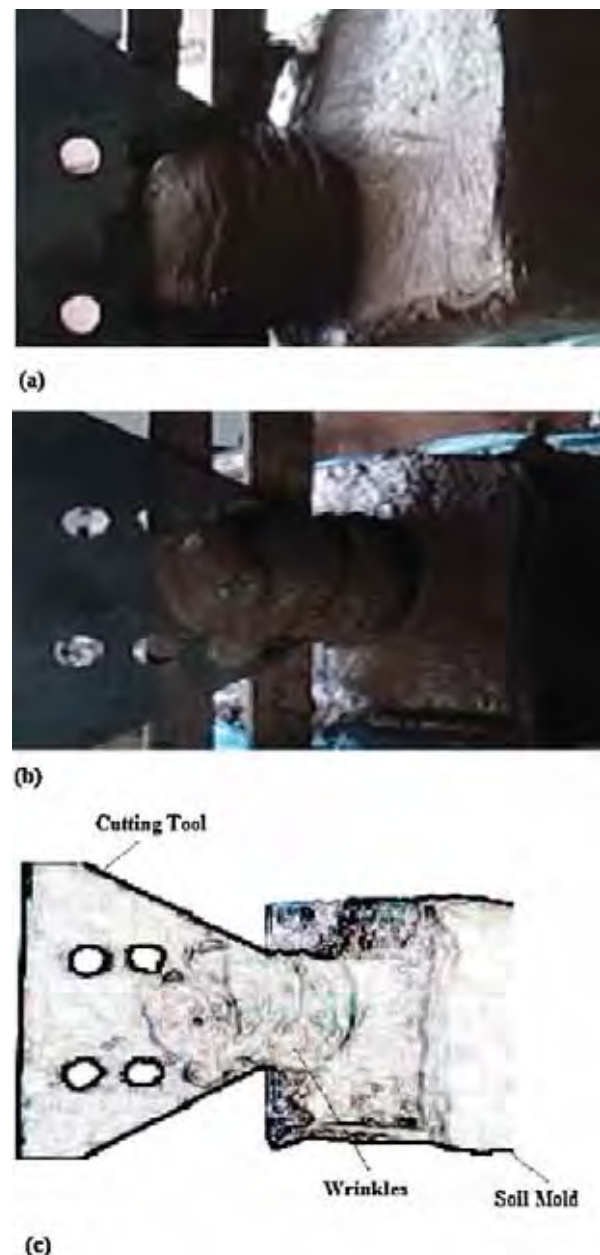


Fig. 6. Flow failure at the liquid limit: (a) dry land soil, (b) paddy soil and (c) schematic sketch.

1996, 2010; Spoor and Godwin, 1978). Makanga et al. (2010) concluded that at 5.2% and 21% moisture content below plastic limit (PL, 23); the force–displacement curves were cyclic in nature and in phase, and matched quite well with the soil deformation characteristics. Rajaram (1987) found cyclic behavior of soil forces with tine movement up to 28% moisture content above plastic limit (PL, 22%). This is consistent with Rajaram and Erbach (1996), who found that in agricultural soils, a tillage tool often causes cyclic failures. The cyclic nature of soil failure patterns was also confirmed by Sharma (1990), Rajaram and Erbach (1998), Makanga (1997), and Jayasuriya (1999), who worked with sandy, clay loam, loamy and lateritic type soils, respectively. In contrast, draft at the liquid limit (Fig. 10) was comparatively diverse and fading. Rajaram and Erbach (1997) concluded that at 40% moisture content (42% saturation), the soil failed by flow and the tine forces were not cyclic. Similarly, Makanga et al. (2010) observed that at





(a)



(b)



(c)

Fig. 7. Flow with considerable bending failure: (a) dry land soil, (b) paddy soil and (c) schematic sketch.

33.5% moisture content (LL, 51%), the force–displacement curves were quite different and had a fading periodicity.

The draft at the sticky, plastic and liquid limits (Figs. 8–10) was slightly higher in paddy soil than in dry land soil, which is

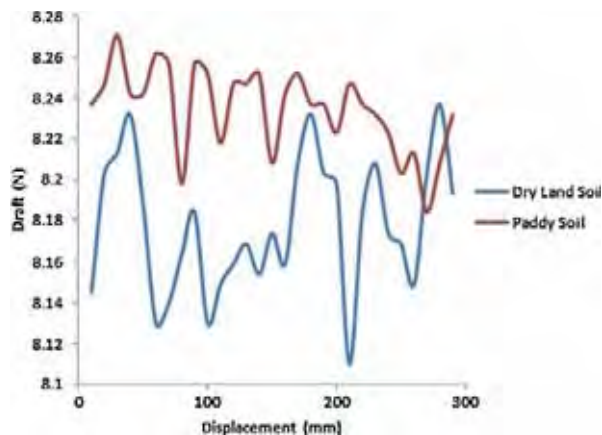


Fig. 8. Draft at the sticky limit in dry land and paddy soils.

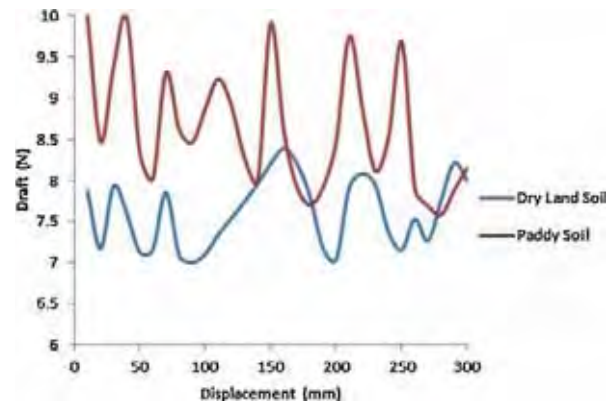


Fig. 9. Draft at the plastic limit in dry land and paddy soils.

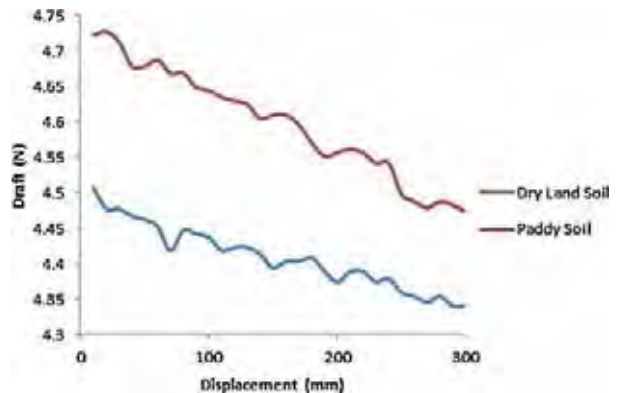


Fig. 10. Draft at the liquid limit in dry land and paddy soils.

attributed to different proportions of silt and clay and soil strengths (Aluko and Seig, 2000). In contrast, the draft at the sticky point (Fig. 11) was identical in both soils due to the high stickiness/adhesion. The highest draft was found at the plastic limit, and the lowest was found at the liquid limit. Although the sticky limit condition was drier than the plastic limit, it had lower draft than the plastic limit. This may be true because the force required to form chip forming failure as well as bending failure at the plastic limit is greater than to break individual aggregates in the case of brittle failure. This is supported by Dexter (1988) and Díaz-Zorita et al. (2002), who found that less energy is required with brittle failure. The variations of peak-to-trough values, wavelength and amplitude were also slightly higher at the plastic limit compared to the sticky limit. At the liquid limit, draft was not

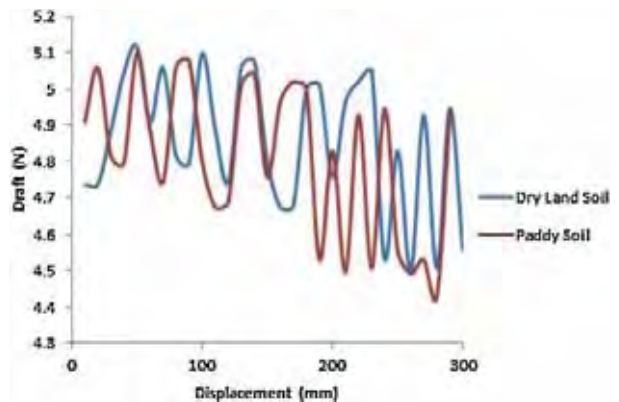


Fig. 11. Draft at the sticky point in dry land and paddy soils.

cyclic; therefore there were a few peak-to-trough values with higher wavelengths. However, the sticky point had the highest variations of peak-to-trough values. This is attributable to the number of soil failures for a given distance of tool travel (Rajaram and Erbach, 1997).

#### 4. Conclusions

This study has shown that there is a direct relationship between soil failure patterns and corresponding draft and the consistency limits of soil. Brittle failure, which is always desirable for good tillage, was obtained at the sticky limit. Chip forming failure was observed at 15° rake angle and 30 mm depth, and bending failure with little strains at 30° and 45° rake angles and 50 mm and 70 mm depths was observed at the plastic limit, while flow failure was linked to the liquid limit of the soil. At the sticky point flow failure was observed at an operating depth of 30 mm and a 15° rake angle, while flow with considerable bending and no strains of elements occurred at 50 mm and 70 mm operating depths and 30° and 45° rake angles. However, bending was more prominent at 70 mm depth and 45° rake angle.

The draft at the sticky limit, plastic limit and sticky point was cyclic in nature, whereas at the liquid limit it was comparatively diverse and fading. The highest draft was found at the plastic limit, and the lowest at the liquid limit. Draft at the sticky limit, plastic limit and liquid limit was higher in paddy soil than in dry land soil, while at the sticky point it was identical. Since the soil failure pattern may change with moisture content, soil type, and particle size distribution within the same textural class, consistency limits of soil can provide clearer and more accurate definitions of soil failure patterns than moisture content levels alone.

#### Acknowledgements

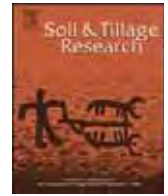
The work was sponsored by the National Science Foundation of China (grant nos. 51275250; 41371238) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). Partial funding was also provided by an NSERC Discovery Grant held by Jan Adamowski.

#### References

- Abo Al-kheer, A., Aoues, Y., Eid, M., El-Hami, A., 2011. Integrating optimization and reliability tools into the design of agricultural machines. In: 20e Congrès Français de Mécanique Besançon, 29 août au 2 septembre 2011. AFM, Maison de la Mécanique, Courbevoie, France. (accessed 29.04.2013) [http://document-sirevues.inist.fr/bitstream/handle/2042/46919/cfm2011\\_327.pdf](http://document-sirevues.inist.fr/bitstream/handle/2042/46919/cfm2011_327.pdf).
- Aluko, O.B., Chandler, H.W., 2004. Characterization and modeling of brittle fracture in two dimensional soil cutting. *Biosyst. Eng.* 88 (3) 369–381, <http://dx.doi.org/10.1016/j.biosystemseng.2004.03.009>.
- Aluko, O.B., Seig, D.A., 2000. An experimental investigation of the characteristics of and conditions for brittle failure in two-dimensional soil cutting. *Soil Till. Res.* 57 (3) 143–157, [http://dx.doi.org/10.1016/S0167-1987\(00\)00156-2](http://dx.doi.org/10.1016/S0167-1987(00)00156-2).
- Archer, J.R., 1975. Soil consistency. In: *Soil Physical Conditions and Crop Production*. Ministry of Agriculture, Fisheries and Food Technical Bulletin 29 His Majesty's Stationery Office, London, pp. 289–297.
- Atterberg, A., 1911. Über die physikalische Bodenuntersuchung und über die Plastizität der Tone. *Int. Mitteilungen für Bodenkunde* 1 (1) 7–9, 10–43.
- Baruah, T.C., Barthakur, H.P., 1997. *A Text Book of Soil Analysis*, 1st ed. Vikas Publishing House, New Delhi.
- Baver, L.D., 1956. *Soil Physics*. John Wiley & Sons, New York.
- Blake, G.R., Hartge, K.H., 1986. Bulk density. In: Klute, A. (Ed.), *Methods of Soil Analysis. Part 1. Physical and Mineralogical Methods*. American Society of Agronomy and Soil Science Society of America, Madison, pp. 363–376.
- Bouyoucos, G.J., 1927. The hydrometer as a new method for the mechanical analysis of soils. *Soil Sci.* 23 (5) 343–353.
- Braja, M.D., 2002. *Principles of Geotechnical Engineering*, 5th ed. Brooks Cole/Thompson Learning, Pacific Grove, CA, pp. 113–114.
- Braunack, M.V., Hewitt, J.S., Dexter, A.R., 1979. Brittle fracture of soil aggregates and the compaction of aggregate beds. *J. Soil Sci.* 30 (4) 653–667, <http://dx.doi.org/10.1111/j.1365-2389.1979.tb01015.x>.
- Campbell, D.J., 2001. Liquid and plastic limits. In: Smith, K.A., Mullins, C. (Eds.), *Soil and Environmental Analysis: Physical Methods*. Marcel Dekker, Inc., New York, pp. 349–376.
- Dexter, A., Bird, N.R.A., 2001. Methods for predicting the optimum and range of soil water contents for tillage based on the water retention curve. *Soil Till. Res.* 57 (4) 203–212, [http://dx.doi.org/10.1016/S0167-1987\(00\)00154-9](http://dx.doi.org/10.1016/S0167-1987(00)00154-9).
- Dexter, A.R., 1988. Advances in characterization of soil structure. *Soil Till. Res.* (11) 199–238.
- Dexter, A.R., 1975. Uniaxial compression of ideal brittle tills. *J. Terramech.* 12 (1) 3–14, [http://dx.doi.org/10.1016/0022-4898\(75\)90003-8](http://dx.doi.org/10.1016/0022-4898(75)90003-8).
- Díaz-Zorita, M., Perfect, E., Grove, J.H., 2002. Disruptive methods for assessing soil structure. *Soil Till. Res.* 64, 3–22.
- Drees, G., 1956. Untersuchungen über das Kräftespiel an Flachbagger-Schneidwerkzeugen im Mittelsand und schwach bindigem, sandigem Schluff: unter besonderer Berücksichtigung der Planierschilde und ebenen Schürfkübel-schneiden Dissertation Aachen Techn Hochschule, Diss., pp. 135.
- Elijah, D.L., Weber, J.A., 1971. Soil failure and pressure patterns for flat cutting blades. *Trans. ASAE* 14 (4) 781–785.
- Fredlund, D.G., Vanapalli, S.K., 2002. Shear strength of unsaturated soils. In: Dane, J.H., Topp, G.C. (Eds.), *Methods of Soil Analysis: Part 4. Physical Methods*, Soil Science Society of America Book Series No. 5. SSSA, Madison, pp. 329–361.
- Godwin, R.J., Spoor, G., 1977. Soil failure with narrow tines. *J. Agric. Eng. Res.* 22 (3) 213–228, [http://dx.doi.org/10.1016/0021-8634\(77\)90044-0](http://dx.doi.org/10.1016/0021-8634(77)90044-0).
- Harison, H.P., 1982. Soil reaction from laboratory studies with an inclined blade. *Trans. ASAE* 25 (1) 7–12.
- Jayasuriya, H.P.W., 1999. *Modelling Soil-Tine Interactions in Lateritic Soils*. (AIT D. Eng. Dissertation, DISS No. AE-99-1) Asian Institute of Technology, Bangkok, Thailand (unpublished).
- Jayasuriya, H.P.W., Salokhe, V.M., 2001. A review of soil-tine models for a range of soil conditions. *J. Agric. Eng. Res.* 79 (1) 1–13, <http://dx.doi.org/10.1006/jaer.2000.0692>.
- Karmakar, S., 2005. *Numerical Modeling of Soil Flow and Pressure Distribution on a Simple Tillage Tool Using Computational Fluid Dynamics*. (Ph.D. Thesis) Dept. Agricultural and Bioresource Engineering, University of Saskatchewan.
- Keller, T., Dexter, A.R., 2012. Plastic limits of agricultural soils as functions of soil texture and organic matter content. *Soil Res.* 50 (1) 7–17, <http://dx.doi.org/10.1071/SR11174>.
- Koolen, A.J., 1972. Mechanical behaviour of soil by treatment with a curved blade having a small angle of approach. *J. Agric. Eng. Res.* 17 (4) 355–367, [http://dx.doi.org/10.1016/S0021-8634\(72\)80043-X](http://dx.doi.org/10.1016/S0021-8634(72)80043-X).
- Koolen, A.J., 1973. Failure patterns in different soils produced by a curved blade with a small angle of approach. N.I.A.E. Subject day on Mechanical Behaviour of Agricultural Soil, Paper 4. National Institute of Agricultural Engineering Report No 7. NIAE, Silsoe, UK.
- Lal, R., Shukla, M.K., 2004. *Principles of Soil Physics*. Marcel Dekker, New York 716 pp.
- Makanga, J.T., 1997. *Soil-Tine Interaction in Loam Soil*. (AIT D. Eng. Dissertation, DISS. No. AE-97-1) Asian Institute of Technology, Bangkok, Thailand (unpublished).
- Makanga, J.T., Salokhe, V.M., Gee-Clough, D., 1996. Effect of tine rake angle and aspect ratio on soil failure patterns in dry loam soil. *J. Terramech.* 33 (5) 233–252, [http://dx.doi.org/10.1016/S0022-4898\(97\)00007-4](http://dx.doi.org/10.1016/S0022-4898(97)00007-4).
- Makanga, J.T., Salokhe, V.M., Gee-Clough, D., 2010. Deformation and force characteristics caused by inclined tines in loam soil with moisture content below liquid limit. *J. Agric. Sci. Technol.* 12 (2) 181–205. (accessed 29.04.2013) <http://elearning.jkuat.ac.ke/journals/ojs/index.php/jagst/article/view/19>.
- McBride, R.A., 2008. Soil consistency: upper and lower plastic limits. In: Carter, M.R., Gregorich, E.G. (Eds.), *Soil Sampling and Methods of Analysis*. 2nd ed. CRC Press, Boca Raton, FL, pp. 761–769.
- Mueller, W., 1985. Standortkundliche Voraussetzungen fuer die Gefuegemellioration durch Tieflockerung im humiden Klima. In: *Die Gefuegemellioration durch Tieflockerung – Bisherige Erfahrungen und Ergebnisse Schriftenreihe des Deutschen Verbandes fuer Wasserwirtschaft und Kulturbau e. V. (DVWK) Heft 70*. Verlag Paul Parey, Hamburg/Berlin, pp. 1–34.
- Müller, L., Schindler, U., 1998a. Wetness criteria for modeling trafficability and workability of cohesive arable soils. In: Brown, L.C. (Ed.), *Proceedings of the Seventh Annual Drainage Symposium on Drainage in the 21st Century: Food Production and Environment*, FL, March 8–10, 1998. ASAE, St. Joseph, Orlando, MI, pp. 472–479.
- Müller, L., Schindler, U., 1998b. Soil moisture and workability of heavy arable soils. *Arch. Agron. Soil Sci.* 44 (2) 161–174, <http://dx.doi.org/10.1080/03650349909366074>.
- Müller, L., Schindler, U., Fausey, N.R., Lal, R., 2003. Comparison of methods for estimating maximum soil water content for optimum workability. *Soil Till. Res.* 72 (1) 9–20, [http://dx.doi.org/10.1016/S0167-1987\(03\)00046-1](http://dx.doi.org/10.1016/S0167-1987(03)00046-1).
- Mullins, C.E., MacLeod, D.A., Northcote, K.H., Tisdall, J.M., Young, I.M., 1990. Hard setting soils: behaviour, occurrence and management. *Adv. Soil Sci.* 11, 37–108, [http://dx.doi.org/10.1007/978-1-4612-3322-0\\_2](http://dx.doi.org/10.1007/978-1-4612-3322-0_2).
- Nalavade, P.P., Salokhe, V.M., Niyamapa, T., Soni, P., 2010. Performance of free rolling and powered tillage discs. *Soil Till. Res.* 109 (2) 87–93, <http://dx.doi.org/10.1016/j.still.2010.05.004>.
- Okunlola, A., Payne, D., 1991. Use of force-deformation curves to estimate Young's modulus and its application to soil aggregate breakdown. *J. Soil Sci.* 42 (4) 543–549, <http://dx.doi.org/10.1111/j.1365-2389.1991.tb00101.x>.
- Rajaram, G., 1987. *Force-Time Behaviour of Tine Implements*. (M. Eng. Thesis No. AE-87-14) Asian Institute of Technology, Bangkok, Thailand (unpublished).

- Rajaram, G., Erbach, D.C., 1996. Soil failure by shear versus soil modification by tillage: a review. *J. Terramech.* 33 (6) 265–272, [http://dx.doi.org/10.1016/S0022-4898\(97\)00009-8](http://dx.doi.org/10.1016/S0022-4898(97)00009-8).
- Rajaram, G., Erbach, D.C., 1997. Hysteresis in soil mechanical behavior. *J. Terramech.* 34 (4) 251–259, [http://dx.doi.org/10.1016/S0022-4898\(98\)00004-4](http://dx.doi.org/10.1016/S0022-4898(98)00004-4).
- Rajaram, G., Erbach, D.C., 1998. Drying stress effect on mechanical behaviour of a clay-loam soil. *Soil Till. Res.* 49 (1–2) 147–158, [http://dx.doi.org/10.1016/S0167-1987\(98\)00163-9](http://dx.doi.org/10.1016/S0167-1987(98)00163-9).
- Rajaram, G., Gee-Clough, D., 1988. Force–distance behavior of tine implements. *J. Agric. Eng. Res.* 41 (2) 81–98, [http://dx.doi.org/10.1016/0021-8634\(88\)90191-6](http://dx.doi.org/10.1016/0021-8634(88)90191-6).
- Salokhe, V.M., 1986. *Cage Wheel Blocking in Wet Clay Soils*. (AIT D. Eng. Dissertation No. AE-86-1) Asian Institute of Technology, Bangkok, Thailand (unpublished).
- Selig, E.T., Nelson, R.D., 1964. Observations of soil cutting with blades. *J. Terramech.* 1 (3) 32–53, [http://dx.doi.org/10.1016/0022-4898\(64\)90038-2](http://dx.doi.org/10.1016/0022-4898(64)90038-2).
- Sharma, V.K., 1990. *Soil–Tool Interactions for Tools of Simple Shape in Dry Sand*. (AIT D. Eng. Dissertation, DISS. No. AE-90-1) Asian Institute of Technology, Bangkok, Thailand (unpublished).
- Smedema, L.K., 1993. Drainage performance and soil management. *Soil Technol.* 6 (2) 183–189, [http://dx.doi.org/10.1016/0933-3630\(93\)90007-2](http://dx.doi.org/10.1016/0933-3630(93)90007-2).
- Smith, C., Johnston, M., Lorentz, S., 1997. Assessing the compaction susceptibility of South African forestry soils. I. The effect of soil type, water content and applied pressure on uni-axial compaction. *Soil Till. Res.* 41, 53–73.
- Snyder, V.A., Miller, R.D., 1989. *Soil deformation and fracture under tensile forces*. In: Larson, W.E., Blake, G.R., Allmaras, R.R., Voorhees, W.B., Gupta, S.C. (Eds.), *Mechanics and Related Processes in Structured Agricultural Soils*, NATO ASI Series E, vol. 172. Kluwer Academic Publishers, Dordrecht, pp. 23–35.
- Sowers, G.F., 1965. Consistency. In: Black, C.A., Evan, D.D., White, J.L., Ensinger, L.E., Clark, F.E. (Eds.), *Methods of Soil Analysis (Part 1), Physical and Mineralogical Properties, Including Statistics of Measurements and Sampling Agronomy Monograph 9*. American Society of Agronomy, Madison, pp. 391–399, <http://dx.doi.org/10.2134/agronmonogr9.1.c31>.
- Spoor, G., Godwin, R.J., 1978. An experimental investigation into the deep loosening of soil by rigid tines. *J. Agric. Eng. Res.* 23 (3) 243–258.
- SSSA, 2009. *Glossary of Soil Science Terms*. SSSA, Madison. (accessed 29.04.2013) <https://www.soils.org/publications/soils-glossary>.
- Stafford, J.V., 1979. The performance of a rigid tine in relation to soil properties and speed. *J. Agric. Eng. Res.* 24 (1) 41–56, [http://dx.doi.org/10.1016/0021-8634\(79\)90059-3](http://dx.doi.org/10.1016/0021-8634(79)90059-3).
- Stafford, J.V., 1981. An application of critical state soil mechanics. The performance of rigid tines. *J. Agric. Eng. Res.* 26 (5) 387–401, [http://dx.doi.org/10.1016/0021-8634\(81\)90115-3](http://dx.doi.org/10.1016/0021-8634(81)90115-3).
- Towner, G.D., 1987. The mechanics of cracking of drying clay. *J. Agric. Eng. Res.* 36 (2) 115–124, [http://dx.doi.org/10.1016/0021-8634\(87\)90118-1](http://dx.doi.org/10.1016/0021-8634(87)90118-1).
- USDA, 1999. *Soil Quality Test Kit Guide*. USDA Soil Quality Institute, Washington, DC.
- Wang, J., 1991. *Stress–Strain Relationship in Wet Clay Soil and Some Applications*. (AIT D. Eng. Dissertation No. AE-91-1) Asian Institute of Technology, Bangkok, Thailand (unpublished).
- Wang, J., Gee-Clough, D., 1993. Deformation and failure in wet clay soil, part 2: soil bin experiments. *J. Agric. Eng. Res.* 54 (1) 57–66, <http://dx.doi.org/10.1006/jaer.1993.1004>.
- Yamamoto, T., 1963. *Soil Moisture Constants and Physical Properties of Selected Soils in Hawaii*. U.S. Forest Service Research Paper PSW-P2. Pacific Southwest Forest and Range Experiment Station, U.S. Forest Service, U.S. Dept. of Agriculture, Berkeley, CA. (accessed 29.04.2013) [http://www.fs.fed.us/psw/publications/documents/psw\\_rp002/psw\\_rp002.pdf](http://www.fs.fed.us/psw/publications/documents/psw_rp002/psw_rp002.pdf).





## Soil failure patterns and draft as influenced by consistency limits: An evaluation of the remolded soil cutting test



A.A. Tagar<sup>a,b</sup>, Changying Ji<sup>a,\*</sup>, Qishuo Ding<sup>a</sup>, Jan Adamowski<sup>c</sup>, F.A. Chandio<sup>a,b</sup>, I.A. Mari<sup>a</sup>

<sup>a</sup> College of Engineering, Nanjing Agricultural University, Nanjing 210031, PR China

<sup>b</sup> Faculty of Agricultural Engineering, Sindh Agriculture University, Tandojam 70060, Pakistan

<sup>c</sup> Department of Bioresource Engineering, McGill University, Ste Anne de Bellevue, Quebec H9X3V9, Canada

### ARTICLE INFO

#### Article history:

Received 2 May 2013

Received in revised form 3 December 2013

Accepted 5 December 2013

#### Keywords:

Sticky limit

Sticky point

Flat triangular shaped tool

Dry land soil

Paddy soil

Laboratory soil bin

### ABSTRACT

Soil failure patterns play an important role in obtaining a better understanding of the mechanical behavior of soils. Despite the large number of studies over the past few decades, a better understanding of soil failure patterns and its relation to soil and tool parameters for particular soils such as dry land and paddy has not been developed. This study investigated soil failure patterns and related draft at sticky, plastic and liquid consistency limits and the sticky point of dry land and paddy soils. A soil cutting test rig was developed to perform soil cutting at three consistency limits (sticky limit, plastic limit and liquid limit) and the sticky point of soil, three rake angles (15°, 30° and 45°), and three operating depths (30 mm, 50 mm and 70 mm). A flat triangular shaped tool operating at a constant speed of 10 mm s<sup>-1</sup> was used in all experiments. Soil failure patterns were observed and recorded using a digital camera, and draft per unit displacement was measured by load cells attached to the soil bin. A direct relationship between soil failure patterns or draft and the consistency limits of soil was found. Brittle failure was observed at the sticky limit, chip forming failure was observed at 15° rake angle and 30 mm depth, and bending failure with little strains of elements at 30° and 45° rake angles and 50 mm and 70 mm depths at plastic limit, while flow failure was linked to the liquid limit of the soil. At the sticky point, flow failure was observed at an operating depth of 30 mm and 15° rake angle, while flow with considerable bending and no strains of elements occurred at 50 mm and 70 mm operating depths and 30° and 45° rake angles. However, bending was more prominent at 70 mm depth and 45° rake angle. The draft at the sticky limit, plastic limit and sticky point was cyclic in nature, whereas at the liquid limit it was comparatively diverse and fading. The highest draft was found at the plastic limit, and the lowest at the liquid limit. Since the soil failure patterns may change with moisture content, soil type and particle size distribution within the same textural class, consistency limits can provide clearer and more accurate definitions of soil failure patterns than moisture content levels alone.

© 2013 Elsevier B.V. All rights reserved.

### 1. Introduction

The proper design and selection of soil-engaging tools to achieve desired soil tillage depends largely on the mechanical behavior of the soils (Rajaram and Erbach, 1998). Soil failure patterns play an important role in obtaining a better understanding of the mechanical behavior of soils under varied soil and tool conditions. The variation in soil failure patterns can be attributed to the variations of mechanical behavior of the soil (Abo Al-kheer et al., 2011). Soil failure patterns can include collapse, fracturing (brittle), chip forming, and flow (Salokhe, 1986; Rajaram and

Gee-Clough, 1988; Sharma, 1990). These failure patterns may vary with soil and tool parameters (Elijah and Weber, 1971; Godwin and Spoor, 1977; Stafford, 1981; Makanga et al., 1996).

Over the past couple of decades, numerous studies have been conducted to evaluate the effects of soil and tool parameters on soil failure patterns such as moisture content (Makanga et al., 2010; Rajaram and Gee-Clough, 1988; Wang and Gee-Clough, 1993), rake angles (Aluko and Seig, 2000), aspect ratio (Makanga et al., 1996) and operating speed (Stafford, 1979; Karmakar, 2005). Despite this large number of studies, a better understanding of the relationships existing between soil failure patterns and soil and tool parameters has not been elucidated. This is particularly true for dry land and paddy soils. This is likely mainly attributable to the complexity of soil genesis, textures, unique weather conditions, and the cropping systems associated with each soil under study.

\* Corresponding author. Tel.: +86 25 58606571.

E-mail addresses: [chyji@njau.edu.cn](mailto:chyji@njau.edu.cn), [chyji@yahoo.com](mailto:chyji@yahoo.com) (C. Ji).



Although the importance of a better understanding of the true failure patterns of soils has been emphasized by a number of authors (e.g., Rajaram and Erbach, 1997, 1998), technical methods to quantify soil failure patterns are limited. Moreover, the numerical value of the moisture content does not show any direct relationship with changes in soil failure patterns in different soils (Jayasuriya and Salokhe, 2001). One potential indicator may relate to a soil's consistency limits. The consistency limits of soil or Atterberg limits (Atterberg, 1911) define the resistance of soils to deformation or rupture. The plastic limit refers to the minimum moisture content at which soil can be puddled (Lal and Shukla, 2004), the liquid limit is the minimum moisture content at which soil flows like a liquid (Dexter and Bird, 2001; SSSA, 2009), and the sticky limit is the moisture content at which soil has little or no stickiness/adhesion to a steel spatula (Lal and Shukla, 2004). These terms are commonly used for classifying cohesive soils for engineering purposes (McBride, 2008). To some extent, they may yield information on the mechanical behavior of soils (Keller and Dexter, 2012). Soil workability is clearly related to the consistency limits of soil (Mueller, 1985; Smedema, 1993; Müller and Schindler, 1998a; Dexter and Bird, 2001; Müller et al., 2003). Archer (1975) concluded that consistency limits can be used to classify soils for workability. Moreover, Yamamoto (1963) suggested that data on the consistency limits of soil can be used as a guide for tillage practices. The sticky point, defined as the moisture content at which soil has maximum stickiness/adhesion to a steel spatula (Braja, 2002), is also an important criteria of soil workability. According to Müller and Schindler (1998b) the stickiness of soil is the limiting factor for its workability.

Many studies have explored soil failure patterns (e.g., Makanga et al., 2010; Wang and Gee-Clough, 1993; Rajaram and Gee-Clough, 1988; Stafford, 1981; Elijah and Weber, 1971). All studies to date have been based on specific soil moisture content levels. Although a few moisture levels could be inferred to represent the agronomically important soil moisture contents at the permanent wilting point, field capacity and saturation (e.g., Makanga et al., 2010), the rationale for choosing specific moisture content values was not specified in these studies. However, some studies provided the plastic limit and liquid limit of their soils, but unfortunately the experimental soil moisture levels that were used did not correspond to moisture content levels at plastic limit or liquid limit. In contrast, our study explores for the first time the use of soil consistency limits to investigate soil failure patterns during soil cutting. In order to provide precise definitions of soil failure patterns, this study was designed to investigate: (i) the relationship between soil failure patterns and draft with consistency limits and (ii) failure patterns and draft at the sticky point, in both dry land and paddy soils.

## 2. Materials and methods

The experiments in this study were carried out in an indoor soil bin test rig developed at the Department of Agricultural

Mechanization, College of Engineering, at Nanjing Agricultural University (NJAU) in China. The soils used in the experiment were dry land soil and paddy soil. These are yellow-brown soils according to the Chinese Soil Taxonomy and Halpudalf in the US classification scheme. China's yellow-brown soils, shown to have arisen through an independent soil genesis mechanism, are distributed across a wide swath of agro-ecological regions (27°33° N lat.), and have important implications for agricultural production. Consequently, an investigation of the general mechanical behaviors of these types of soil during cutting is warranted, particularly with respect to clearly defining its failure patterns. The paddy soil was used for a rice-wheat rotation on the university's Jiangpu Experimental Farm, while the farm's dry land soil was used to cultivate vegetables such as potatoes (*Solanum tuberosum* L., tomatoes (*Solanum lycopersicum* L.), eggplants (*Solanum melongena* L.) and chilies (*Capsicum* sp.).

### 2.1. Soil preparation

The experimental soil was air dried for two to three weeks, ground, and sieved through a 4 mm sieve. Composite soil samples were taken from the sieved soil to determine the moisture content in the soil, and then on the basis of existing moisture content a calculated amount of water was added to the soil (Eq. (1)) as a fine spray so as to attain the desired moisture content at the required consistency limit and the sticky point of the soil.

$$V_a = V_{req} - V_{ex} = (SMC_{req} \times W_s) - (SMC_{ex} \times W_s) \quad (1)$$

where,  $V_{req}$  is the volume of water which needs to be added to the soil in order to achieve the desired soil moisture content (ml),  $V_{ex}$  is the existing water present initially in the dry sieved soil (ml),  $SMC_{req}$  is the required moisture content ( $\text{g kg}^{-1}$ ),  $SMC_{ex}$  is the moisture content of the dry sieved soil ( $\text{g kg}^{-1}$ ), and  $W_s$  is the weight of soil (g).

It was then mixed well to obtain a homogenous soil specimen, covered with a polyethylene sheet to prevent evaporation, and left for 24 h to equilibrate to uniform moisture content. The soil was transferred to a metal-framed mold and compacted to the ideal bulk densities i.e., 1.22–1.4  $\text{mg m}^{-3}$  for sandy clay loam soil (USDA, 1999) using a manually operated compactor. If this was not done, soil compaction could have had effects on soil failure patterns. Soil molds at the sticky limit were compacted to 1.22–1.25  $\text{mg m}^{-3}$ . However, soil molds at the plastic limit, liquid limit and sticky point were compacted to 1.3–1.4  $\text{mg m}^{-3}$  (Table 1). Because of high moisture content, it was not possible to compact soil molds at the required density with the compaction device. This is consistent with Smith et al. (1997), who concluded that when the moisture content is so high that all the soil pores are filled with water, the soil becomes less compressible. These soil molds (300 mm × 100 mm × 100 mm) were then transferred to the soil cutting test rig.

**Table 1**

The physical and mechanical properties of dry land and paddy soils at different consistency limits and the sticky point.

Consistency limit	Dry land soil					Paddy soil				
	Moisture content ( $\text{g kg}^{-1}$ )	Dry bulk density ( $\text{mg m}^{-3}$ )	Cohesion (kPa)	Angle of internal friction (Deg.)	Cone index (kPa)	Moisture content ( $\text{g kg}^{-1}$ )	Dry bulk density ( $\text{mg m}^{-3}$ )	Cohesion (kPa)	Angle of internal friction (Deg.)	Cone index (kPa)
Sticky limit	15	1.22	0.0012	26	106	17.7	1.25	0.001	29	109
Plastic limit	22	1.32	0.0005	16	104	32	1.4	0.0004	18	106
Liquid limit	36	1.3	0	4	11	45	1.4	0	4	13
Sticky point	32	1.4	0.00005	8	18	44	1.3	0	4	15

**Table 2**

The physical and mechanical properties of dry land and paddy soils after different failure patterns.

Consistency limit	Type of failure pattern	Dry land soil				Paddy soil			
		Dry bulk density ( $\text{mg m}^{-3}$ )	Cohesion (kPa)	Angle of internal friction (Deg.)	Cone index (kPa)	Dry bulk density ( $\text{mg m}^{-3}$ )	Cohesion (kPa)	Angle of internal friction (Deg.)	Cone index (kPa)
Sticky limit	Brittle failure	1.06	0.0006	20.5	38.78	1.08	0.0004	26	39.33
Plastic limit	Chip forming	1.33	0.0004	15.5	36	1.32	0.0003	19	34.78
	Bending Failure	1.39	0.0006	16	42	1.37	0.0005	19	40
Liquid limit	Flow failure	1.3	0	5.5	4.44	1.31	0	4.5	8.22
Sticky point	Flow and bending	1.39	0	5.2	6	1.31	0	4.6	6

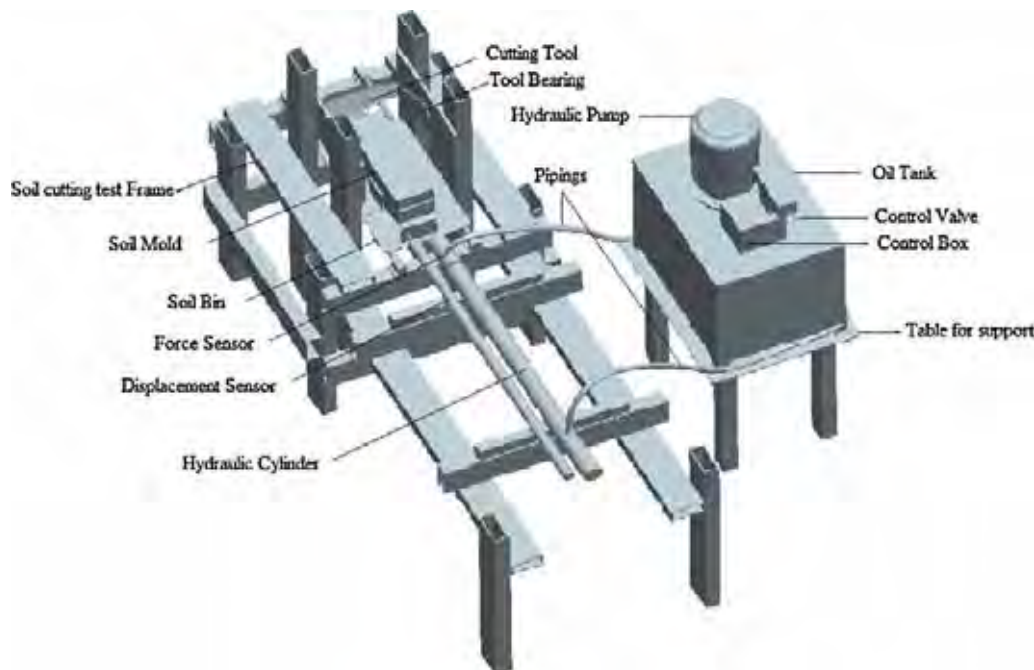
## 2.2. Soil physical and mechanical properties

Soil texture was determined by the Bouyoucos Hydrometer Method (Bouyoucos, 1927), soil moisture content and dry bulk density by the gravimetric method (Blake and Hartge, 1986), shear strength properties using a direct shear box apparatus (Fredlund and Vanapalli, 2002). The cone index (penetration resistance) was directly measured with a digital soil compaction meter (TJSD-750, Zhejiang Top Instrument Co. Ltd, China). The dry land soil was composed of 67% sand, 28% clay and 5% silt, and was classified as sandy clay loam, while the paddy soil was comprised of 50% sand, 24% clay, and 26% silt, and classified as sandy clay loam according to the USDA textural classification. The sticky limit of soil was measured as gravimetric moisture content at which soil fails to stick or adhere to a stainless steel spatula when the spatula blade is drawn across the face of the moist unknaded mass of soil, exerting a firm pressure against the soil (Baruah and Barthakur, 1997). The plastic limit was determined as the gravimetric moisture content at which a soil just begins to crumble as it is rolled into a thread of 3 mm diameter (Sowers, 1965). A 30° cone mounted on a shaft, with a total weight of about 80 g, was allowed to drop on a cup full of soil for 5 s. Soil moisture content corresponding to a penetration of 20 mm on the linear relationship between soil moisture content ( $x$ -axis) and the penetration ( $y$ -axis) was determined and considered as the cone penetrometer liquid limit

(Campbell, 2001). The sticky point of soil was measured at the point at which the soil had maximum stickiness/adhesion to steel spatula (Baver, 1956). The gravimetric soil moisture content at the respective consistency limits and sticky point differed between the two soils. The organic carbon content of dry land soil was  $8.4 \text{ g kg}^{-1}$ , and for paddy soil it was  $9.6 \text{ g kg}^{-1}$ . The physical and mechanical properties of dry land and paddy soils before the cutting test and after different failure patterns are shown in Tables 1 and 2.

## 2.3. Soil cutting test rig

A safe and easy to operate test rig was developed for the study. Its main features include a soil bin, tool bearing, cutting tool, hydraulic system, load cells, and a computer based data acquisition and control system. The rectangular steel tubings ( $90 \text{ mm} \times 40 \text{ mm} \times 5 \text{ mm}$ ), were cut into two sizes (i.e.,  $8 \text{ mm} \times 1000 \text{ mm}$  and  $16 \text{ mm} \times 600 \text{ mm}$ ), and were joined through bolts to construct a 2000 mm long and 600 mm wide test rig (Fig. 1). Two railings were provided on the test rig; a soil bin (500 mm long and 300 mm wide) with two supports ( $90 \text{ mm} \times 40 \text{ mm} \times 50 \text{ mm}$ ) on the right side and one at the front was mounted on the railings and fixed through steel screws. The height of the supports was kept smaller to avoid boundary frictions. Two steel sheets ( $200 \text{ mm} \times 150 \text{ mm} \times 10 \text{ mm}$ ) were fixed in the test rig at a distance of

**Fig. 1.** Soil cutting test rig used in the experiments.

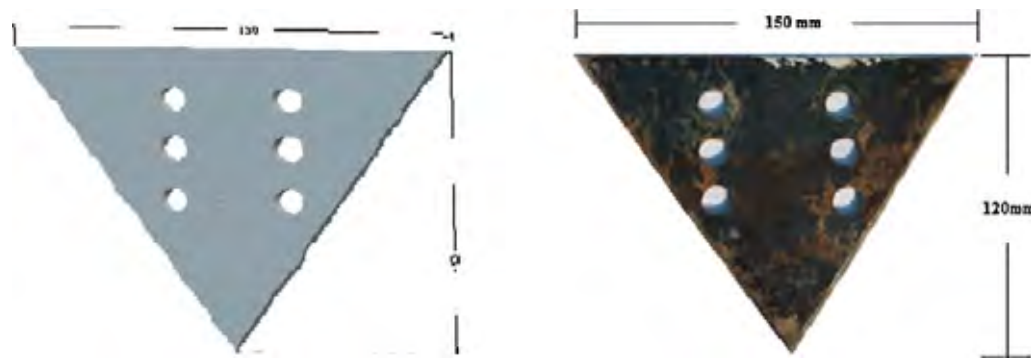


Fig. 2. Cutting tool used in the experiments.

100 mm on the left and right side of the soil bin, and the tool bearing was attached to them. The holes (15 mm diameter) were drilled into steel sheets at different angles so that their rake angles could be adjusted according to the conditions tested as done by Nalavade et al. (2010), Aluko and Seig (2000), and Rajaram and Erbach (1998). In order to move the soil bin to perform cutting operations, the soil bin was connected to the hydraulic system powered by an electric motor (2.2 kW) through a hydraulic cylinder. A two way solenoid valve was provided to move the soil bin (i) to perform cutting operations and (ii) to take it back to its original position. Two load cells connected to a computer were attached to the soil bin and a program was developed using LabVIEW (National Instruments, Austin, TX, USA) for data acquisition. The load cell designated as LSR-2A (2KN, Shanghai, Zhendan Sensor and Instrument Factory, China) was fixed in the hydraulic cylinder and attached to the soil bin to measure draft, whereas the load cell designated as W-DCD400 (1000 mm, Hefei, Gaochuang Sensor Co. Ltd. China) was placed near the hydraulic cylinder on the test rig and was attached to the soil bin to measure displacement. These load cells were calibrated prior to the experiments, similarly as described by Aluko and Seig (2000), and Makanga et al. (2010).

#### 2.4. Soil cutting test procedure

The soil cutting test was performed at three consistency limits (sticky limit, plastic limit and liquid limit) and at the sticky point of the soil, factorially combined with three rake angles (15°, 30° and 45°) and three operating depths (30 mm, 50 mm and 70 mm), as done by Wang (1991), Makanga et al. (1996), and Aluko and Seig (2000). A flat triangular shaped tool operating at a constant speed of  $10 \text{ mm s}^{-1}$  was used in all experiments to allow for a clear observation and recording of soil failure patterns (Makanga et al., 1996, 2010). In order to adjust the cutting tool at different depths, three rows of holes (15 mm diameter), two in each row with 20 mm spacing, were drilled about 20 mm from the top at the center of the cutting tool (Fig. 2).

The soil molds were placed in the soil bin. As the cutting tool moved in the soil bin, the resulting soil failure patterns were observed and recorded using a digital camera (Canon Power Shot A4000 IS: 16 Mega pixels, Canon Inc., China). The recorded videos were then converted to snapshots. Simultaneously, the force-displacement signals were recorded through a data acquisition system (High Speed Advantech Bus Multifunction Data Acquisition Card USB-4711A with a resolution of 12 bits per channel sampling rate up to 150 ks/s) in an output file in the computer loaded with LabVIEW. This was then converted to MS-Excel spreadsheets to calculate draft along with the distance traveled per unit millimeter of 300 mm soil mold.

### 3. Results and discussion

#### 3.1. Soil failure patterns at different consistency limits and the sticky point

Soil failed in a brittle failure pattern in both dry land and paddy soils at all operating depths (30 mm, 50 mm and 70 mm) and rake angles (15°, 30° and 45°) at the sticky limit of soil (Fig. 3). However, in dry land soil the size of clod separated from the soil mold was 116 mm long, 49 mm wide and 100 mm thick; in the paddy soil the size of clod separated from the soil mold was 209 mm long, 62 mm wide and 100 mm thick. The above results showed a direct relationship between brittle failure at the sticky limit of both dry land and paddy soils. The close relation of the brittle failure pattern with the sticky limit demonstrates that, if brittle failure is desired (Dexter, 1975; Towner, 1987; Braunack et al., 1979; Snyder and Miller, 1989; Mullins et al., 1990; Okunlola and Payne, 1991; Aluko and Seig, 2000), soil should be ploughed at its sticky limit. Although this type of failure has been reported in a number of studies (Drees, 1956; Selig and Nelson, 1964; Elijah and Weber, 1971; Koolen, 1972, 1973; Rajaram and Gee-Clough, 1988; Rajaram and Erbach, 1998; Aluko and Seig, 2000; Aluko and Chandler, 2004), the soil conditions in those studies were not properly described.

Rajaram and Erbach (1996) pointed out that brittle failure or cracking of the soil in front of the tool occurs in moist soils, with moisture contents not exceeding the plastic limit. Similarly, Jayasuriya and Salokhe (2001) stated that brittle failure occurs in frictional-cohesive and cohesive soils when the moisture content is closer to (below) the plastic limit. However, the moist conditions closer to (below) are also difficult to recognize. According to Rajaram and Gee-Clough (1988), brittle failure occurs at 18.3% moisture content below plastic limit (PL, 22%). In contrast, Stafford (1981) obtained brittle failure pattern at 18% and 28% moisture content below and above the plastic limit (PL, 21%). Moreover, Makanga et al. (2010) found progressive shear type failure patterns at 21% moisture content below plastic limit (PL, 23%). Similarly, Harison (1982) observed progressive shear type failure at 18–20% moisture contents. In our study, brittle failure was found at the sticky limit of the soil.

A chip-forming failure pattern was found at 15° rake angle and 30 mm depth at the plastic limit in both dry land and paddy soils. The chip was 43 mm wide and 30 mm thick in dry land soil, while in paddy soil the chip was 62 mm wide and 30 mm thick. Rajaram (1987) and Rajaram and Gee-Clough (1988) observed chip formation failure at 28.6% moisture content above the plastic limit (PL, 22%) or between the plastic limit and saturation with vertical tines of varying width (30, 120 and 150 mm) at 100 mm depth. The bending failure with little strains of elements in the vertical direction was witnessed at 30° and 45° rake angles and



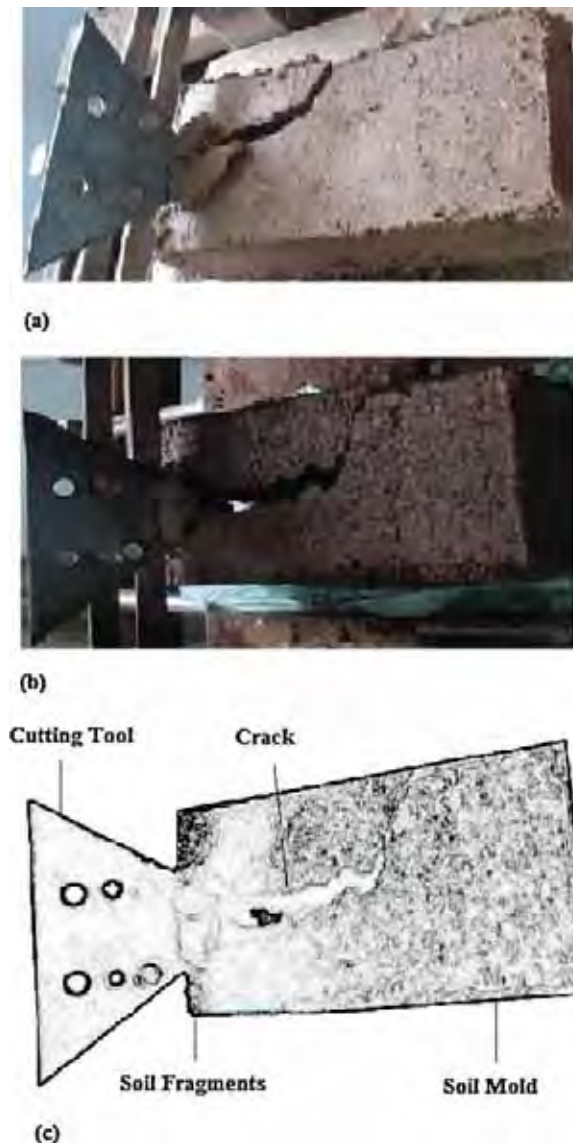


Fig. 3. Brittle failure at the sticky limit: (a) dry land soil, (b) paddy soil and (c) schematic sketch.

50 mm and 70 mm depths (Figs. 4 and 5). In dry land soil, the bending failure was 125 mm wide and 56 mm thick, while in paddy soil the bending failure was 106 mm wide and 53 mm thick. Elijah and Weber (1971) observed bending failure with inclined flat blades at 150 mm depth and a 45° rake angle above the plastic limit. At 70 mm depth and 45° rake angle, bending failure occurred in a reverse or backward direction. This indicated that at the plastic limit, chip forming failure may be observed at <50 mm depths and <30° rake angles, while bending failure with little strains of elements in the vertical direction occurred at ≥50 mm operating depths and ≥30° rake angles.

At the liquid limit, flow failure was observed in both soils at all operating depths (30 mm, 50 mm and 70 mm) and rake angles (15°, 30° and 45°) as shown in Fig. 6. Flow failure was 57 mm wide and 33 mm thick in dry land soil, while in paddy soil it was 43 mm wide and 32 mm thick. This type of failure pattern was also observed by Wang (1991) who conducted an experiment with a wide tine with varying rake angles (25–125°) in a Bangkok clay soil at 44% and 52% soil moisture content below the liquid limit (LL, 54%). He observed flow type soil failure at both soil moisture contents. Similarly, Rajaram and Gee-Clough (1988) observed

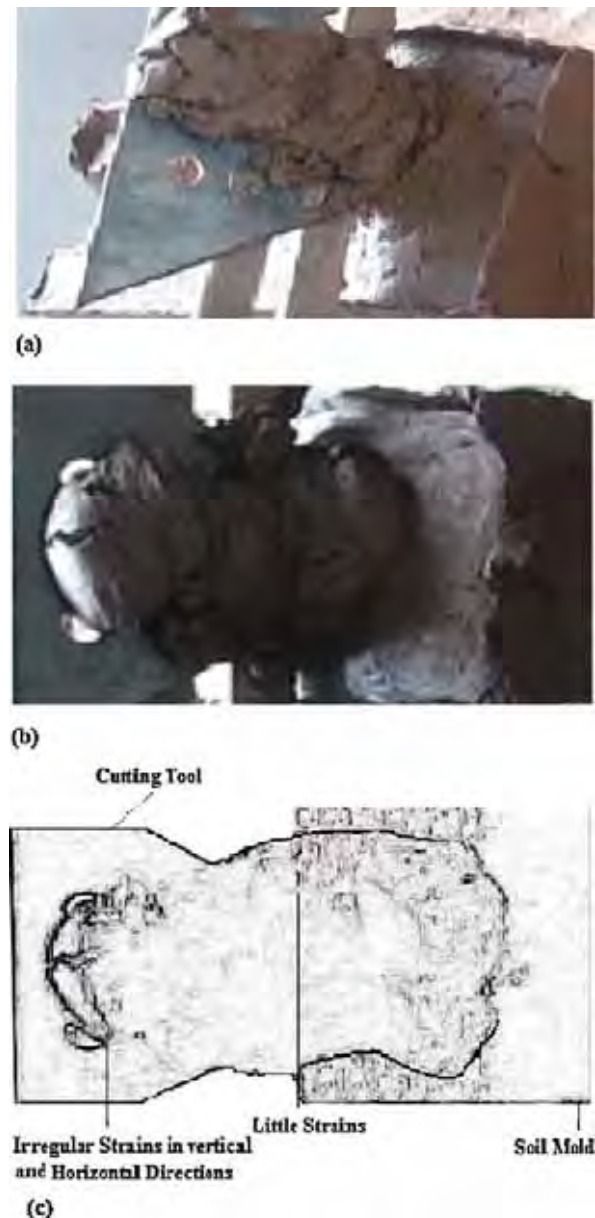


Fig. 4. Chip forming failure at the plastic limit: (a) dry land soil, (b) paddy soil and (c) schematic sketch.

flow-type soil failure at 42% below the liquid limit (LL, 46%). In contrast, Wang and Gee-Clough (1993), in studying a Bangkok clay soil (wet), found that at a soil moisture content as high as 44%, soil brittle failure was found at small and medium rake angles (<50°), as well as soil shear failure pattern. However, our study found that flow failure is clearly related to the liquid limit of the soil.

At the sticky point flow failure was observed at 30 mm depth and 15° rake angle, while flow with considerable bending and no strains of elements was experienced at 50 mm and 70 mm operating depths and 30° and 45° rake angles, as can be seen from Fig. 7. In dry land soil, the flow with considerable bending failure was 100.5 mm wide and 57 mm thick, while in paddy soil the flow with considerable bending failure was 108 mm wide and 55 mm thick. However, bending was more prominent at 70 mm depth and 45° rake angle, possibly because during cutting at 30 mm depth and 15° rake angle, the soil in contact with the cutting tool is less than compared to that of 50 mm and 70 mm operating depths and 30° and 45° rake angles. It is anticipated that





Fig. 5. Bending failure at the plastic limit: (a) dry land soil, (b) paddy soil and (c) schematic sketch.

if soil is ploughed at the sticky point at  $>70$  mm operating depths and  $>45^\circ$  rake angles, and at a speed of  $10 \text{ mm s}^{-1}$ , bending will be high enough to stop further soil cutting, because of the high stickiness and greater amount of soil that is in contact with the cutting tool.

Although flow failure (Stafford, 1981; Rajaram and Gee-Clough, 1988; Wang, 1991; Makanga et al., 1996, 2010) and bending failure patterns (Elijah and Weber, 1971) have been explored before in the literature, flow failure with considerable bending and no strains of elements at the sticky point are reported for the first time in the present study.

### 3.2. Draft at different consistency limits and the sticky point

The draft at the sticky limit (Fig. 8), plastic limit (Fig. 9) and sticky point (Fig. 11) were cyclic in nature. The cyclic nature could be related to the observed repetitive and periodic nature of the soil failure patterns at different stages of tool travel (Makanga et al.,

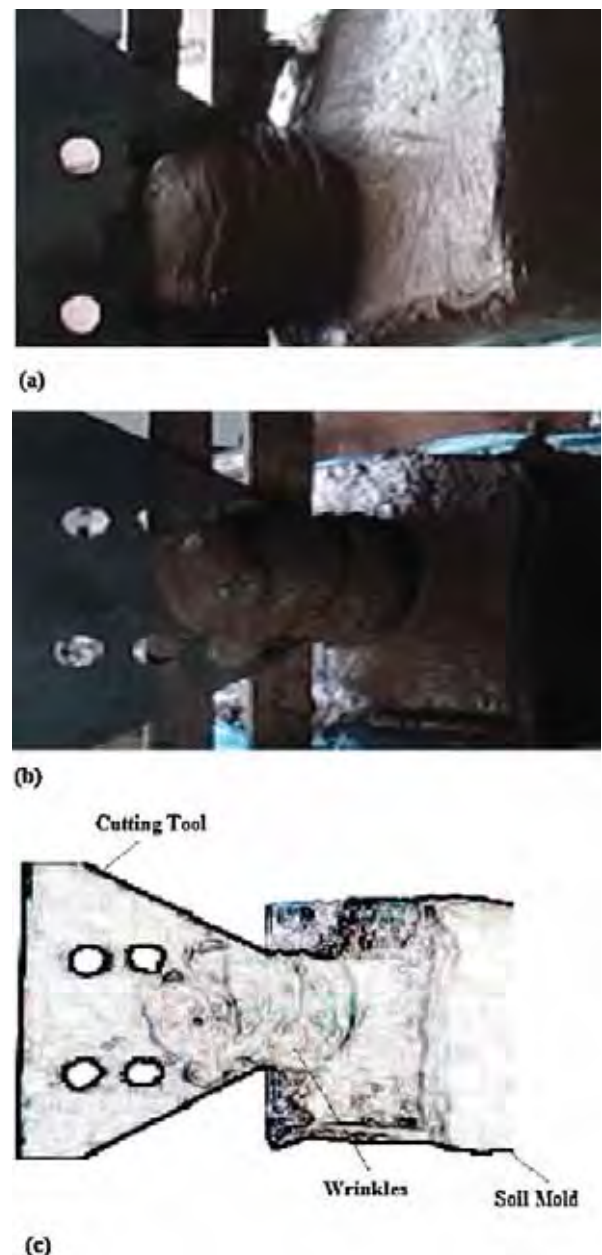


Fig. 6. Flow failure at the liquid limit: (a) dry land soil, (b) paddy soil and (c) schematic sketch.

1996, 2010; Spoor and Godwin, 1978). Makanga et al. (2010) concluded that at 5.2% and 21% moisture content below plastic limit (PL, 23); the force–displacement curves were cyclic in nature and in phase, and matched quite well with the soil deformation characteristics. Rajaram (1987) found cyclic behavior of soil forces with tine movement up to 28% moisture content above plastic limit (PL, 22%). This is consistent with Rajaram and Erbach (1996), who found that in agricultural soils, a tillage tool often causes cyclic failures. The cyclic nature of soil failure patterns was also confirmed by Sharma (1990), Rajaram and Erbach (1998), Makanga (1997), and Jayasuriya (1999), who worked with sandy, clay loam, loamy and lateritic type soils, respectively. In contrast, draft at the liquid limit (Fig. 10) was comparatively diverse and fading. Rajaram and Erbach (1997) concluded that at 40% moisture content (42% saturation), the soil failed by flow and the tine forces were not cyclic. Similarly, Makanga et al. (2010) observed that at

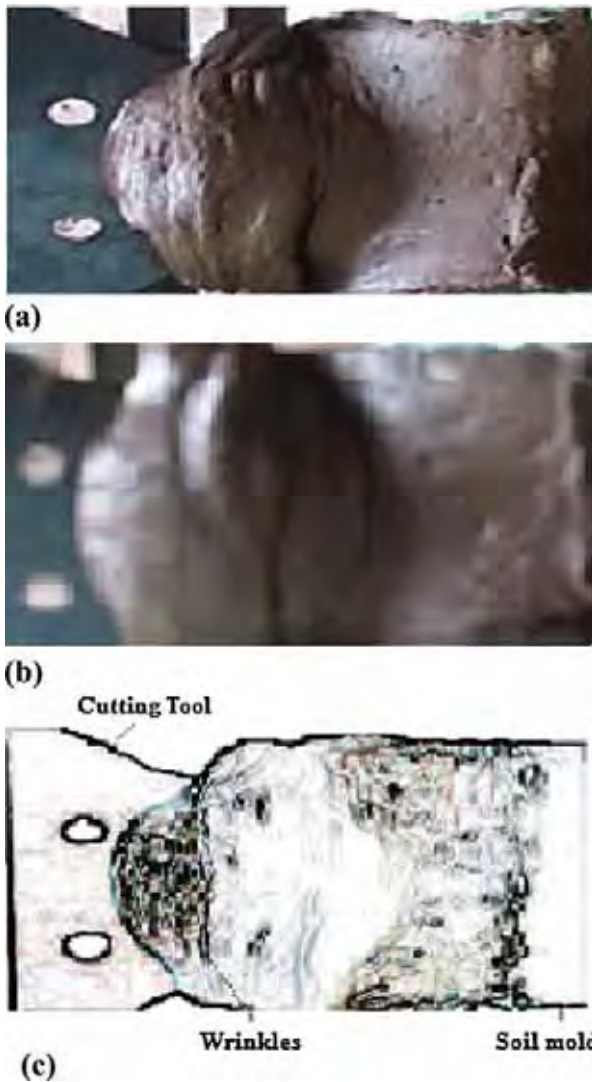


Fig. 7. Flow with considerable bending failure: (a) dry land soil, (b) paddy soil and (c) schematic sketch.

33.5% moisture content (LL, 51%), the force–displacement curves were quite different and had a fading periodicity.

The draft at the sticky, plastic and liquid limits (Figs. 8–10) was slightly higher in paddy soil than in dry land soil, which is

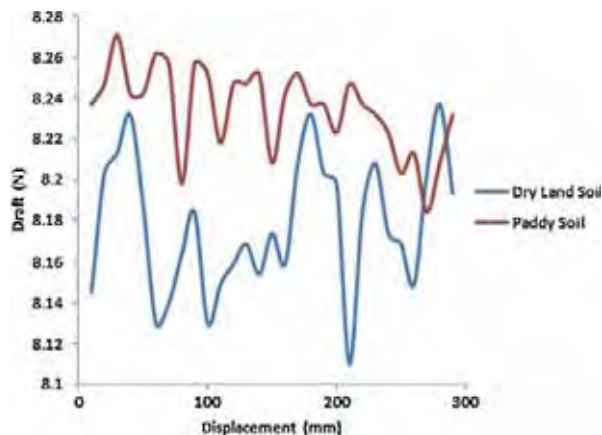


Fig. 8. Draft at the sticky limit in dry land and paddy soils.

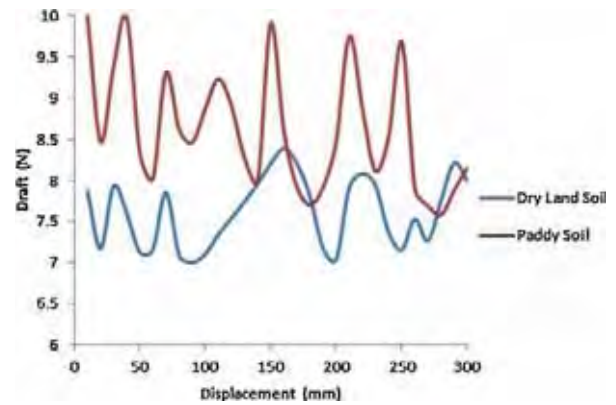


Fig. 9. Draft at the plastic limit in dry land and paddy soils.

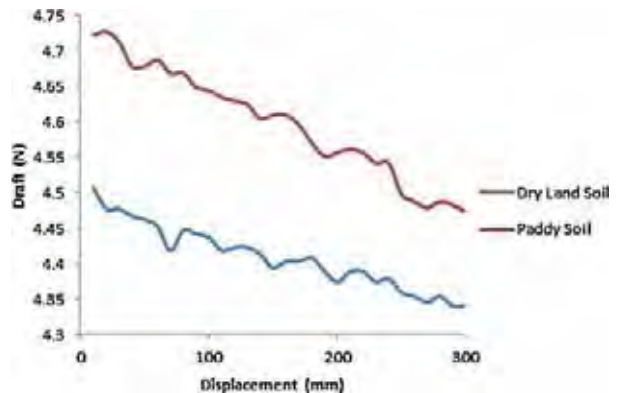


Fig. 10. Draft at the liquid limit in dry land and paddy soils.

attributed to different proportions of silt and clay and soil strengths (Aluko and Seig, 2000). In contrast, the draft at the sticky point (Fig. 11) was identical in both soils due to the high stickiness/adhesion. The highest draft was found at the plastic limit, and the lowest was found at the liquid limit. Although the sticky limit condition was drier than the plastic limit, it had lower draft than the plastic limit. This may be true because the force required to form chip forming failure as well as bending failure at the plastic limit is greater than to break individual aggregates in the case of brittle failure. This is supported by Dexter (1988) and Díaz-Zorita et al. (2002), who found that less energy is required with brittle failure. The variations of peak-to-trough values, wavelength and amplitude were also slightly higher at the plastic limit compared to the sticky limit. At the liquid limit, draft was not

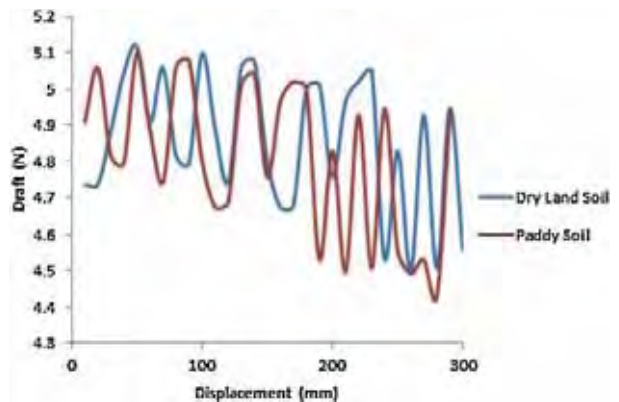


Fig. 11. Draft at the sticky point in dry land and paddy soils.



cyclic; therefore there were a few peak-to-trough values with higher wavelengths. However, the sticky point had the highest variations of peak-to-trough values. This is attributable to the number of soil failures for a given distance of tool travel (Rajaram and Erbach, 1997).

#### 4. Conclusions

This study has shown that there is a direct relationship between soil failure patterns and corresponding draft and the consistency limits of soil. Brittle failure, which is always desirable for good tillage, was obtained at the sticky limit. Chip forming failure was observed at 15° rake angle and 30 mm depth, and bending failure with little strains at 30° and 45° rake angles and 50 mm and 70 mm depths was observed at the plastic limit, while flow failure was linked to the liquid limit of the soil. At the sticky point flow failure was observed at an operating depth of 30 mm and a 15° rake angle, while flow with considerable bending and no strains of elements occurred at 50 mm and 70 mm operating depths and 30° and 45° rake angles. However, bending was more prominent at 70 mm depth and 45° rake angle.

The draft at the sticky limit, plastic limit and sticky point was cyclic in nature, whereas at the liquid limit it was comparatively diverse and fading. The highest draft was found at the plastic limit, and the lowest at the liquid limit. Draft at the sticky limit, plastic limit and liquid limit was higher in paddy soil than in dry land soil, while at the sticky point it was identical. Since the soil failure pattern may change with moisture content, soil type, and particle size distribution within the same textural class, consistency limits of soil can provide clearer and more accurate definitions of soil failure patterns than moisture content levels alone.

#### Acknowledgements

The work was sponsored by the National Science Foundation of China (grant nos. 51275250; 41371238) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). Partial funding was also provided by an NSERC Discovery Grant held by Jan Adamowski.

#### References

- Abo Al-kheer, A., Aoues, Y., Eid, M., El-Hami, A., 2011. Integrating optimization and reliability tools into the design of agricultural machines. In: 20e Congrès Français de Mécanique Besançon, 29 août au 2 septembre 2011. AFM, Maison de la Mécanique, Courbevoie, France. (accessed 29.04.2013) [http://document-sirevues.inist.fr/bitstream/handle/2042/46919/cfm2011\\_327.pdf](http://document-sirevues.inist.fr/bitstream/handle/2042/46919/cfm2011_327.pdf).
- Aluko, O.B., Chandler, H.W., 2004. Characterization and modeling of brittle fracture in two dimensional soil cutting. *Biosyst. Eng.* 88 (3) 369–381, <http://dx.doi.org/10.1016/j.biosystemseng.2004.03.009>.
- Aluko, O.B., Seig, D.A., 2000. An experimental investigation of the characteristics of and conditions for brittle failure in two-dimensional soil cutting. *Soil Till. Res.* 57 (3) 143–157, [http://dx.doi.org/10.1016/S0167-1987\(00\)00156-2](http://dx.doi.org/10.1016/S0167-1987(00)00156-2).
- Archer, J.R., 1975. Soil consistency. In: *Soil Physical Conditions and Crop Production*. Ministry of Agriculture, Fisheries and Food Technical Bulletin 29 His Majesty's Stationery Office, London, pp. 289–297.
- Atterberg, A., 1911. Über die physikalische Bodenuntersuchung und über die Plastizität der Tone. *Int. Mitteilungen für Bodenkunde* 1 (1) 7–9, 10–43.
- Baruah, T.C., Barthakur, H.P., 1997. *A Text Book of Soil Analysis*, 1st ed. Vikas Publishing House, New Delhi.
- Baver, L.D., 1956. *Soil Physics*. John Wiley & Sons, New York.
- Blake, G.R., Hartge, K.H., 1986. Bulk density. In: Klute, A. (Ed.), *Methods of Soil Analysis. Part 1. Physical and Mineralogical Methods*. American Society of Agronomy and Soil Science Society of America, Madison, pp. 363–376.
- Bouyoucos, G.J., 1927. The hydrometer as a new method for the mechanical analysis of soils. *Soil Sci.* 23 (5) 343–353.
- Braja, M.D., 2002. *Principles of Geotechnical Engineering*, 5th ed. Brooks Cole/Thompson Learning, Pacific Grove, CA, pp. 113–114.
- Braunack, M.V., Hewitt, J.S., Dexter, A.R., 1979. Brittle fracture of soil aggregates and the compaction of aggregate beds. *J. Soil Sci.* 30 (4) 653–667, <http://dx.doi.org/10.1111/j.1365-2389.1979.tb01015.x>.
- Campbell, D.J., 2001. Liquid and plastic limits. In: Smith, K.A., Mullins, C. (Eds.), *Soil and Environmental Analysis: Physical Methods*. Marcel Dekker, Inc., New York, pp. 349–376.
- Dexter, A., Bird, N.R.A., 2001. Methods for predicting the optimum and range of soil water contents for tillage based on the water retention curve. *Soil Till. Res.* 57 (4) 203–212, [http://dx.doi.org/10.1016/S0167-1987\(00\)00154-9](http://dx.doi.org/10.1016/S0167-1987(00)00154-9).
- Dexter, A.R., 1988. Advances in characterization of soil structure. *Soil Till. Res.* (11) 199–238.
- Dexter, A.R., 1975. Uniaxial compression of ideal brittle tills. *J. Terramech.* 12 (1) 3–14, [http://dx.doi.org/10.1016/0022-4898\(75\)90003-8](http://dx.doi.org/10.1016/0022-4898(75)90003-8).
- Díaz-Zorita, M., Perfect, E., Grove, J.H., 2002. Disruptive methods for assessing soil structure. *Soil Till. Res.* 64, 3–22.
- Drees, G., 1956. Untersuchungen über das Kräftespiel an Flachbagger-Schneidwerkzeugen im Mittelsand und schwach bindigem, sandigem Schluff: unter besonderer Berücksichtigung der Planierschilde und ebenen Schürfkübelnschneiden. *Dissertation Aachen Techn Hochschule*, Diss., pp. 135.
- Elijah, D.L., Weber, J.A., 1971. Soil failure and pressure patterns for flat cutting blades. *Trans. ASAE* 14 (4) 781–785.
- Fredlund, D.G., Vanapalli, S.K., 2002. Shear strength of unsaturated soils. In: Dane, J.H., Topp, G.C. (Eds.), *Methods of Soil Analysis: Part 4. Physical Methods*, Soil Science Society of America Book Series No. 5. SSSA, Madison, pp. 329–361.
- Godwin, R.J., Spoor, G., 1977. Soil failure with narrow tines. *J. Agric. Eng. Res.* 22 (3) 213–228, [http://dx.doi.org/10.1016/0021-8634\(77\)90044-0](http://dx.doi.org/10.1016/0021-8634(77)90044-0).
- Harison, H.P., 1982. Soil reaction from laboratory studies with an inclined blade. *Trans. ASAE* 25 (1) 7–12.
- Jayasuriya, H.P.W., 1999. *Modelling Soil-Tine Interactions in Lateritic Soils*. (AIT D. Eng. Dissertation, DISS No. AE-99-1) Asian Institute of Technology, Bangkok, Thailand (unpublished).
- Jayasuriya, H.P.W., Salokhe, V.M., 2001. A review of soil-tine models for a range of soil conditions. *J. Agric. Eng. Res.* 79 (1) 1–13, <http://dx.doi.org/10.1006/jaer.2000.0692>.
- Karmakar, S., 2005. *Numerical Modeling of Soil Flow and Pressure Distribution on a Simple Tillage Tool Using Computational Fluid Dynamics*. (Ph.D. Thesis) Dept. Agricultural and Bioresource Engineering, University of Saskatchewan.
- Keller, T., Dexter, A.R., 2012. Plastic limits of agricultural soils as functions of soil texture and organic matter content. *Soil Res.* 50 (1) 7–17, <http://dx.doi.org/10.1071/SR11174>.
- Koolen, A.J., 1972. Mechanical behaviour of soil by treatment with a curved blade having a small angle of approach. *J. Agric. Eng. Res.* 17 (4) 355–367, [http://dx.doi.org/10.1016/S0021-8634\(72\)80043-X](http://dx.doi.org/10.1016/S0021-8634(72)80043-X).
- Koolen, A.J., 1973. Failure patterns in different soils produced by a curved blade with a small angle of approach. N.I.A.E. Subject day on Mechanical Behaviour of Agricultural Soil, Paper 4. National Institute of Agricultural Engineering Report No 7. NIAE, Silsoe, UK.
- Lal, R., Shukla, M.K., 2004. *Principles of Soil Physics*. Marcel Dekker, New York 716 pp.
- Makanga, J.T., 1997. *Soil-Tine Interaction in Loam Soil*. (AIT D. Eng. Dissertation, DISS. No. AE-97-1) Asian Institute of Technology, Bangkok, Thailand (unpublished).
- Makanga, J.T., Salokhe, V.M., Gee-Clough, D., 1996. Effect of tine rake angle and aspect ratio on soil failure patterns in dry loam soil. *J. Terramech.* 33 (5) 233–252, [http://dx.doi.org/10.1016/S0022-4898\(97\)00007-4](http://dx.doi.org/10.1016/S0022-4898(97)00007-4).
- Makanga, J.T., Salokhe, V.M., Gee-Clough, D., 2010. Deformation and force characteristics caused by inclined tines in loam soil with moisture content below liquid limit. *J. Agric. Sci. Technol.* 12 (2) 181–205. (accessed 29.04.2013) <http://elearning.jkuat.ac.ke/journals/ojs/index.php/jagst/article/view/19>.
- McBride, R.A., 2008. Soil consistency: upper and lower plastic limits. In: Carter, M.R., Gregorich, E.G. (Eds.), *Soil Sampling and Methods of Analysis*. 2nd ed. CRC Press, Boca Raton, FL, pp. 761–769.
- Mueller, W., 1985. Standortkundliche Voraussetzungen fuer die Gefuegemellioration durch Tieflockerung im humiden Klima. In: *Die Gefuegemellioration durch Tieflockerung – Bisherige Erfahrungen und Ergebnisse Schriftenreihe des Deutschen Verbandes fuer Wasserwirtschaft und Kulturbau e. V. (DVWK) Heft 70*. Verlag Paul Parey, Hamburg/Berlin, pp. 1–34.
- Müller, L., Schindler, U., 1998a. Wetness criteria for modeling trafficability and workability of cohesive arable soils. In: Brown, L.C. (Ed.), *Proceedings of the Seventh Annual Drainage Symposium on Drainage in the 21st Century: Food Production and Environment*, FL, March 8–10, 1998. ASAE, St. Joseph, Orlando, MI, pp. 472–479.
- Müller, L., Schindler, U., 1998b. Soil moisture and workability of heavy arable soils. *Arch. Agron. Soil Sci.* 44 (2) 161–174, <http://dx.doi.org/10.1080/03650349909366074>.
- Müller, L., Schindler, U., Fausey, N.R., Lal, R., 2003. Comparison of methods for estimating maximum soil water content for optimum workability. *Soil Till. Res.* 72 (1) 9–20, [http://dx.doi.org/10.1016/S0167-1987\(03\)00046-1](http://dx.doi.org/10.1016/S0167-1987(03)00046-1).
- Mullins, C.E., MacLeod, D.A., Northcote, K.H., Tisdall, J.M., Young, I.M., 1990. Hard setting soils: behaviour, occurrence and management. *Adv. Soil Sci.* 11, 37–108, [http://dx.doi.org/10.1007/978-1-4612-3322-0\\_2](http://dx.doi.org/10.1007/978-1-4612-3322-0_2).
- Nalavade, P.P., Salokhe, V.M., Niyamapa, T., Soni, P., 2010. Performance of free rolling and powered tillage discs. *Soil Till. Res.* 109 (2) 87–93, <http://dx.doi.org/10.1016/j.still.2010.05.004>.
- Okunlola, A., Payne, D., 1991. Use of force-deformation curves to estimate Young's modulus and its application to soil aggregate breakdown. *J. Soil Sci.* 42 (4) 543–549, <http://dx.doi.org/10.1111/j.1365-2389.1991.tb00101.x>.
- Rajaram, G., 1987. *Force-Time Behaviour of Tine Implements*. (M. Eng. Thesis No. AE-87-14) Asian Institute of Technology, Bangkok, Thailand (unpublished).

- Rajaram, G., Erbach, D.C., 1996. Soil failure by shear versus soil modification by tillage: a review. *J. Terramech.* 33 (6) 265–272, [http://dx.doi.org/10.1016/S0022-4898\(97\)00009-8](http://dx.doi.org/10.1016/S0022-4898(97)00009-8).
- Rajaram, G., Erbach, D.C., 1997. Hysteresis in soil mechanical behavior. *J. Terramech.* 34 (4) 251–259, [http://dx.doi.org/10.1016/S0022-4898\(98\)00004-4](http://dx.doi.org/10.1016/S0022-4898(98)00004-4).
- Rajaram, G., Erbach, D.C., 1998. Drying stress effect on mechanical behaviour of a clay-loam soil. *Soil Till. Res.* 49 (1–2) 147–158, [http://dx.doi.org/10.1016/S0167-1987\(98\)00163-9](http://dx.doi.org/10.1016/S0167-1987(98)00163-9).
- Rajaram, G., Gee-Clough, D., 1988. Force–distance behavior of tine implements. *J. Agric. Eng. Res.* 41 (2) 81–98, [http://dx.doi.org/10.1016/0021-8634\(88\)90191-6](http://dx.doi.org/10.1016/0021-8634(88)90191-6).
- Salokhe, V.M., 1986. *Cage Wheel Blocking in Wet Clay Soils*. (AIT D. Eng. Dissertation No. AE-86-1) Asian Institute of Technology, Bangkok, Thailand (unpublished).
- Selig, E.T., Nelson, R.D., 1964. Observations of soil cutting with blades. *J. Terramech.* 1 (3) 32–53, [http://dx.doi.org/10.1016/0022-4898\(64\)90038-2](http://dx.doi.org/10.1016/0022-4898(64)90038-2).
- Sharma, V.K., 1990. *Soil–Tool Interactions for Tools of Simple Shape in Dry Sand*. (AIT D. Eng. Dissertation, DISS. No. AE-90-1) Asian Institute of Technology, Bangkok, Thailand (unpublished).
- Smedema, L.K., 1993. Drainage performance and soil management. *Soil Technol.* 6 (2) 183–189, [http://dx.doi.org/10.1016/0933-3630\(93\)90007-2](http://dx.doi.org/10.1016/0933-3630(93)90007-2).
- Smith, C., Johnston, M., Lorentz, S., 1997. Assessing the compaction susceptibility of South African forestry soils. I. The effect of soil type, water content and applied pressure on uni-axial compaction. *Soil Till. Res.* 41, 53–73.
- Snyder, V.A., Miller, R.D., 1989. *Soil deformation and fracture under tensile forces*. In: Larson, W.E., Blake, G.R., Allmaras, R.R., Voorhees, W.B., Gupta, S.C. (Eds.), *Mechanics and Related Processes in Structured Agricultural Soils*, NATO ASI Series E, vol. 172. Kluwer Academic Publishers, Dordrecht, pp. 23–35.
- Sowers, G.F., 1965. Consistency. In: Black, C.A., Evan, D.D., White, J.L., Ensinger, L.E., Clark, F.E. (Eds.), *Methods of Soil Analysis (Part 1), Physical and Mineralogical Properties, Including Statistics of Measurements and Sampling Agronomy Monograph 9*. American Society of Agronomy, Madison, pp. 391–399, <http://dx.doi.org/10.2134/agronmonogr9.1.c31>.
- Spoor, G., Godwin, R.J., 1978. An experimental investigation into the deep loosening of soil by rigid tines. *J. Agric. Eng. Res.* 23 (3) 243–258.
- SSSA, 2009. *Glossary of Soil Science Terms*. SSSA, Madison. (accessed 29.04.2013) <https://www.soils.org/publications/soils-glossary>.
- Stafford, J.V., 1979. The performance of a rigid tine in relation to soil properties and speed. *J. Agric. Eng. Res.* 24 (1) 41–56, [http://dx.doi.org/10.1016/0021-8634\(79\)90059-3](http://dx.doi.org/10.1016/0021-8634(79)90059-3).
- Stafford, J.V., 1981. An application of critical state soil mechanics. The performance of rigid tines. *J. Agric. Eng. Res.* 26 (5) 387–401, [http://dx.doi.org/10.1016/0021-8634\(81\)90115-3](http://dx.doi.org/10.1016/0021-8634(81)90115-3).
- Towner, G.D., 1987. The mechanics of cracking of drying clay. *J. Agric. Eng. Res.* 36 (2) 115–124, [http://dx.doi.org/10.1016/0021-8634\(87\)90118-1](http://dx.doi.org/10.1016/0021-8634(87)90118-1).
- USDA, 1999. *Soil Quality Test Kit Guide*. USDA Soil Quality Institute, Washington, DC.
- Wang, J., 1991. *Stress–Strain Relationship in Wet Clay Soil and Some Applications*. (AIT D. Eng. Dissertation No. AE-91-1) Asian Institute of Technology, Bangkok, Thailand (unpublished).
- Wang, J., Gee-Clough, D., 1993. Deformation and failure in wet clay soil, part 2: soil bin experiments. *J. Agric. Eng. Res.* 54 (1) 57–66, <http://dx.doi.org/10.1006/jaer.1993.1004>.
- Yamamoto, T., 1963. *Soil Moisture Constants and Physical Properties of Selected Soils in Hawaii*. U.S. Forest Service Research Paper PSW-P2. Pacific Southwest Forest and Range Experiment Station, U.S. Forest Service, U.S. Dept. of Agriculture, Berkeley, CA. (accessed 29.04.2013) [http://www.fs.fed.us/psw/publications/documents/psw\\_rp002/psw\\_rp002.pdf](http://www.fs.fed.us/psw/publications/documents/psw_rp002/psw_rp002.pdf).





# Trend detection in surface air temperature in Ontario and Quebec, Canada during 1967–2006 using the discrete wavelet transform



D. Nalley<sup>a</sup>, J. Adamowski<sup>a,\*</sup>, B. Khalil<sup>a</sup>, B. Ozga-Zielinski<sup>b</sup>

<sup>a</sup> Department of Bioresource Engineering, Faculty of Agriculture and Environmental Sciences, McGill University, Ste-Anne-de-Bellevue, Quebec H9X 3V9, Canada

<sup>b</sup> Department of Environment Protection and Development, Environmental Engineering Faculty, Warsaw University of Technology, ul. Nowowiejska 20, Warsaw 00-653, Poland

## ARTICLE INFO

### Article history:

Received 3 April 2013

Received in revised form 29 June 2013

Accepted 30 June 2013

### Keywords:

Temperature

Trend

Climate change

Discrete wavelet transform

Mann–Kendall trend test

Canada

## ABSTRACT

The main purpose of this study is to detect trends in the mean surface air temperature over the southern parts of Ontario and Quebec, Canada, for the period of 1967–2006. This is accomplished by determining the most dominant periodic components that affect trends in different temperature data categories (monthly, seasonally-based, seasonal, and annual), which were obtained from a total of five stations. The discrete wavelet transform (DWT) technique, the Mann–Kendall (MK) trend test, and sequential Mann–Kendall analysis were used in this study – co-utilizing these techniques in temperature trend studies has not been explored extensively. The mother wavelet, number of decomposition levels, and boundary condition were determined using a newly proposed criterion based on the relative error of the MK Z-values between the original data and the approximation component of the last decomposition level. This study found that all stations experienced positive trends: significant trends were observed in all of the monthly, seasonally-based, and annual data. For the different seasons, although the trend values were all positive, not all stations experienced significant trends. It was found that high-frequency components ranging from 2 to 12 months were more prominent for trends in the higher resolution data (i.e. monthly and seasonally based). The positive trends observed for the annual data are thought to be mostly attributable to warming during winter and summer seasons, which are manifested in the form of multiyear to decadal events (mostly between 8 and 16 years).

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

The impacts of climate variability can be assessed by analyzing trends in surface air temperature. According to the latest assessment report by the IPCC (IPCC, 2007), mean global surface air temperature has experienced an increase of between 0.56 °C and 0.92 °C for the period from 1906 to 2005. Changes in surface air temperature as a result of

changing climate have serious ramifications on the hydrological cycle (and therefore, on water resources) and the surface energy budget (Vincent et al., 2007). Examples of these consequences are: intensification of the hydrological cycle (Mishra and Singh, 2010), modification of hydrological indicators such as seasonal runoff, precipitation, streamflow, and potential evapotranspiration (Mimikou et al., 2000; Labat et al., 2004), more severe flood discharges (Ludwig et al., 2004), sea-level rise (which has serious implications on the economy and societies in general) (Nicholls and Tol, 2006), and increased risks of health-related problems (Karaburun et al., 2011.), etc.

Since climate change is directly linked to temperature, a large number of studies have been undertaken globally and

\* Corresponding author at: Department of Bioresource Engineering, Faculty of Agricultural and Environmental Sciences, McGill University, Ste-Anne-de-Bellevue, Quebec H9X3V9, Canada. Tel.: +1 514 224 0492.

E-mail addresses: [deasy.nalley@mail.mcgill.ca](mailto:deasy.nalley@mail.mcgill.ca) (D. Nalley), [jan.adamowski@mcgill.ca](mailto:jan.adamowski@mcgill.ca) (J. Adamowski), [bahaa.e.khalil@mail.mcgill.ca](mailto:bahaa.e.khalil@mail.mcgill.ca) (B. Khalil).

regionally in order to assess temperature trends and to quantify the impacts of increasing temperature. Most of these studies have found the existence of positive trends in different temperature indices. For example, Shrestha et al. (1999) found that the temperature patterns in Nepal were increasing after 1977, with a rate ranging from 0.03 to 0.12 °C per year. Domroes and El-Tantawi (2005) reported increasing mean temperature trends in northern Egypt over the period 1941–2000. Fan and Wang (2011) studied climate change by looking at the monotonic trends in annual and seasonal air temperature indices across Shanxi province in China, and found that there have been warming trends in temperature over the period 1959–2008. Karaburun et al. (2011) also analyzed the spatio-temporal patterns of temperature change in Istanbul, Turkey for the period of 1975 to 2006, and observed that warmer temperature trends generally prevailed for seasonal and annual temperature indices. Although many studies have found that all seasons showed positive trends, winter usually experiences the greatest warming (e.g. Lund et al., 2001; Rebetz and Reinhard, 2007; Fan and Wang, 2011). Fan and Wang (2011) observed that winter warming in Shanxi province in China was statistically significant at less than the 0.1% significance level. Rebetz and Reinhard (2007) also found that the greatest warming occurred during winter (1975–2004) in Switzerland.

It has been mentioned that the northern hemisphere has been experiencing more temperature warming since the 1950s compared to the southern hemispheres (Jones and Moberg, 2003; IPCC, 2007; Chaouche et al., 2010; Karaburun et al., 2011). More specifically, North America is expected to experience a warmer climate, in which the increase in the mean annual temperature could be more than that of the global mean increase (IPCC, 2007). Moberg et al. (2005) reconstructed long-term proxy data from tree-ring and sediments in lakes and oceans to analyze the variation in temperature in the northern hemisphere. They found that there has never been any period within the past 2000 years that is as warm as post 1990 (Moberg et al., 2005). Lund et al. (2001) and Lu et al. (2005) also found that there were increasing temperature trends during 1922–2000 in the East, West-Coast, and northern Midwest of the USA. All seasons in the contiguous 48 states experienced increasing temperature trends with winter showing the highest warming (Lu et al., 2005).

Zhang et al. (2000) mentioned that it may be easier to assess climate change in countries such as Canada because according to Nicholls et al. (1996), human-induced climate change is foreseen to be more severe in high-latitude areas. As such, numerous studies on trends of Canadian temperature indices have been conducted both at the national and regional levels. Zhang et al. (2000) provided comprehensive information on Canada's temperature and precipitation trends. At the national level it was found that the annual mean temperature has experienced an increase of approximately 1 °C during the last half of the 20th century (Zhang et al., 2000). For areas below 60°N, Zhang et al. (2000) showed that although the trends are not monotonic during 1900–1998 and differed from region to region, there is a statistically significant increase in the mean annual temperature that was caused by the increases before the 1940s and after the 1970s. Mohsin and Gough (2010) analyzed the temperature trends in a smaller spatial scale covering Toronto and the Greater Toronto area for the period 1970–2000. The trend analysis on the annual minimum

and mean temperature indices showed that urbanization has contributed to the observed warming trend experienced by the urban stations (Toronto Pearson exhibited the highest warming trend). Similarly, Prokoph and Patterson (2004) and Adamowski and Prokoph (2013) found that the temperature in urban Ottawa was going up by more than 0.01 °C per year compared to the nearby rural areas over the past 100 years, which was associated with the population growth and urban heat island effects.

Studies investigating trends in temperature commonly involve the use of the Mann–Kendall (MK) trend test. This trend test is usually preferred over other statistical tests because of its robustness and power. The MK test may be used even if the analyzed data does not follow a Gaussian normal frequency distribution (Kadioğlu, 1997). Chaouche et al. (2010) chose to employ the MK test in studying climatic indices (including temperature) in the context of climate change because trends are assumed to be slowly changing phenomena; they also mentioned that even if a break change occurs, the MK test is still powerful. Having said this, the use of the original MK test only gives accurate results when the test is applied on data that are free from serial correlation (Mohsin and Gough, 2010). Hirsch and Slack (1984) and Hamed and Rao (1998) proposed modifications to the original MK test in order to account for seasonality and serial correlation factors that may be present in a time series.

When analyzing temperature time series, it is important to examine the behavior of the different low- and high-frequency components contained in the data, which represent fluctuations such as inter-annual and decadal events (Baliunas et al., 1997). Detecting and identifying trends in non-stationary temperature datasets is complicated due to factors such as changes in climate that occur in non-monotonic and non-uniform fashions. Additionally, variability and trends that are observed in a dataset can be associated with climate noise (Franzke, 2010). Trends in temperature can also be enhanced or reduced by the variability in the temperature itself (that may occur in the form of different time scales) and changes in climate that occur naturally. Prior to computing a trend assessment, it is important to define what a trend is. Since many climate data exhibit nonlinear and non-stationary characteristics, commonly defined trends using a simple straight line fit on to the data are not suitable (Wu et al., 2007). Analyzing temperature trends that involves the use of methods such as moving average, regression analysis, and Fourier-based techniques may not also be appropriate because of their assumptions that involve stationary and linear inferences (Wu et al., 2007).

A spectral analysis method that has been found to be very useful for analyzing geophysical time series (which are often characterized by non-stationarity) is the wavelet transform (WT) (Lau and Weng, 1995; Lindsay et al., 1996). The WT is suitable for decomposing one-dimensional non-stationary time series into its two-dimensional (time-frequency) information (Lau and Weng, 1995; Torrence and Compo, 1998; Pišoft et al., 2004). In this study, the non-linear trends are assumed to have occurred in a gradual manner and are contained in the low-frequency components of the data.

Numerous studies have acknowledged that the WT is superior for use in analyzing non-stationary data compared to conventional spectral analysis methods, such as the Fourier transform (FT). The FT decomposes signals into sine wave

functions, which have unlimited duration, whereas the WT – having an irregular and non-symmetrical function shape – decomposes signals into wavelet functions that have limited duration and zero mean (Drago and Boxall, 2002). The WT was used to separate the high-frequency components and low-frequency components of the temperature data used in this study. The last decomposition level contains the lowest-frequency component of the analyzed temperature data, which represents the trend component of the data. The different decomposition levels – which represent different periodic time scales – were analyzed for trends. The goal was to then determine the most important periodic components that affect the observed MK temperature trends. This was accomplished by using the MK test (and its sequential version) to evaluate the behavior of the different periodic components resulting from the DWT which range from intra-annual to decadal time scales. Lower periodic components may be related to factors such as seasonality, which can be reflected in the form of semi-annual and annual cycles. Large-scale climate circulations, some of which tend to occur in a long time-scale and are thus associated with higher decomposition levels or periodic components. For example, if 11-year or 22-year periodic oscillations are present and found to be important in the spectrum of a climate data analysis, they may be related to the Sun's activity cycles: the 11-year sunspot cycle and the 22-year solar magnetic cycle.

The use of WT has been seen in several climatological applications, including studies of spatiotemporal patterns of temperature changes. Baliunas et al. (1997) employed the WT in order to analyze temperature trends and their time-scale information in central England. The 7.5-year scale was observed to be the most stable peak within the 2- to 105-year time-scale range (Baliunas et al., 1997). Pišoft et al. (2004) were able to demonstrate that the global wavelet spectrum of the continuous wavelet transform (CWT) had better accuracy in determining the years of local maxima for the longest time periods of the studied Czech temperature series compared to the FT – the wavelet power was able to reveal different features and activities of each periodic component. Similarly, Jung et al. (2002) and Prokoph and Patterson (2004) utilized the CWT to analyze the warming trends in the winter temperature data in South Korea and Ottawa, respectively. The results of the WA for the winter temperature over South Korea revealed that the decadal and inter-decadal events of 16 and 33 years were strongly persistent during 1954–1999 (Jung et al., 2002). A slightly weaker inter-annual event of 4.9 years was found to be associated with the El Niño cycle (Jung et al., 2002). The results of CWT in studying urban warming trends in Ottawa, Ontario revealed that multi-decadal and inter-seasonal periodic modes are thought to contribute to the winter warming in Ottawa, which is related to the urban heat island (Prokoph and Patterson, 2004; Adamowski and Prokoph, 2013).

The majority of studies that employed WT in investigating trends in temperature used the CWT approach (e.g. Baliunas et al., 1997; Jung et al., 2002; Polyakov et al., 2003; Pišoft et al., 2004; Prokoph and Patterson, 2004; and Kravchenko et al., 2011). This is due to the fact that the CWT allows for the analysis of data at all locations of time and space (Wang and Lu, 2009). However, rather than producing a one-dimensional time series, CWT produces a two-dimensional scalogram, which may contain redundant information (Percival, 2008).

Furthermore, edge effects associated with the application of the CWT complicate signal reconstructions (Adamowski et al., 2009). If the discrete wavelet transform (DWT) approach is chosen, the decomposition process is simplified but still efficient because the computation is based on a dyadic discretization (integer powers of two) (Chou, 2007). This generates a compact representation of the analyzed signal (Wang and Lu, 2009) and thus, the redundancy of the information is reduced. Achieving perfect signal reconstructions is also relatively simple when the DWT approach is used.

The main purpose of this study is to analyze trends in four temperature categories – monthly, seasonally-based, seasonal, and annual – by combining the use of the DWT approach with the MK trend test. The DWT is used to decompose the time series into their different lower-resolution components; the MK test was applied to each time series resulting from the decomposition in order to assess their statistical significance. The dyadic arrangement used in the DWT procedure allowed us to investigate the contributions of periodic events – ranging from 2 months to 32 years – to the observed trends over the 40-year study period. Although there are many temperature trend detection studies that have been conducted in Canada, these studies have not explored the contribution of the high- and low-frequency components of the analyzed data to the observed trends using the DWT and the MK trend test. Additionally, temperature trend studies that focus on localized areas in Canada are still relatively rare.

## 2. Theoretical background

### 2.1. Time-scale representation of signals by the wavelet transform (WT)

The wavelet transform is a mathematical tool that uses wave functions – known as wavelets – similar to sine and cosine functions. A wavelet must satisfy the admissibility condition of having a zero mean (Farge, 1992; Torrence and Compo, 1998). The property of the WT in which it is localized in time and frequency domains is very useful because it allows for the extraction of the different modes of variability that vary in time (Lim and Lye, 2004). The window used in the WT can be adjusted to the whole time-frequency domain – it can be dilated and shifted with a resolution that is adjustable in both time and frequency domains (Lau and Weng, 1995). The narrow and wide windows are used to capture the high-frequency and low-frequency components of the signal, respectively (Lau and Weng, 1995). Therefore, when analyzing a signal, WT separates the signal's high frequency (short periodic components) and low frequency (long periodic components) constituents (Drago and Boxall, 2002). This is one of the main reasons that the WT is more advantageous when used for decomposing signals with non-stationary characteristics, compared to the more conventional spectral analysis, such as the Fourier transform (FT) or windowed Fourier transform (WFT). Fourier Transform uses sine and cosine functions, which do not account for the time information of the signals being analyzed. It therefore cannot provide how information has changed from one time interval to the next (Lau and Weng, 1995). With the WFT, the window used to analyze a time series is fixed, so when there are many different frequencies involved in

the time series, the fixed window picks up more of the high-frequency information but little low-frequency information (Lau and Weng, 1995).

In order to decompose a time series using the WT, the mother wavelet is translated along the signal in a number of steps (using high-pass and low-pass filters). This procedure then produces wavelet coefficients, which measure the correlation of the wavelet to the original signal at a specific scale as a function of time – this is the time-scale representation of the signal, which holds information about the magnitude and location of different events at difference scales (Lindsay et al., 1996; Drago and Boxall, 2002). Different scales are represented by different stretched versions of the mother wavelet.

The WT can be performed either in continuous or discrete modes. Signal reconstructions from the wavelet coefficients are relatively simple to compute when the DWT approach is used – this is done by using the inverse filter function of the wavelet transform (Torrence and Compo, 1998). Signal reconstructions for the CWT, on the other hand, are somewhat problematic because of the redundancy in the time-scale information (Torrence and Compo, 1998). The DWT mode also operates on dyadic scales separating the analyzed signal scale by scale – this is another advantage of using the DWT approach (Lindsay et al., 1996). The signal decomposition using DWT starts out with the smallest scales and continues to larger scales, doubling in size for each round of operation.

The decomposition of a time series  $x_t$ , via the WT is accomplished using the following function (Lau and Weng, 1995):

$$\Psi_{a,s}(t) = \frac{1}{\sqrt{s}} \Psi\left(\frac{t-a}{s}\right) \tag{1}$$

where  $s$  (which is greater than zero) represents the scaling factor,  $a$  is the translation factor, and  $\Psi(t)$  is the analyzing wavelet. The wavelet coefficients ( $C$ ) via the CWT for the time series  $x_t$  (with equal time interval,  $dt$ ), are calculated as follows (Lau and Weng, 1995):

$$C(a,s) = \frac{1}{\sqrt{s}} \int \Psi^*\left(\frac{t-a}{s}\right) x(t) dt \tag{2}$$

where  $\Psi^*$  is the complex conjugate number based on the scaling ( $s$ ) and translation ( $a$ ) factors. The wavelet coefficients ( $w$ ) via the discrete wavelet approach for the time series (with dyadic grid arrangement) are calculated as follows (Partal and Küçük, 2006):

$$w(a,s) = \frac{1}{(2)^{\frac{t}{2}}} \sum_{t=0}^{N-1} x_t \Psi\left(\frac{t}{2^s} - a\right). \tag{3}$$

### 2.2. The original Mann–Kendall (MK) trend test

The original MK test is based on Mann (1945) and Kendall (1975). It is a rank correlation test for two sets of observations between the rank order of the recorded values and their ordered values in time. The null hypothesis of the MK test for a dataset ( $X_h, h = 1, 2, 3, \dots, n$ ) is that the dataset is independent and identically distributed (Yue et al., 2002). The alternative hypothesis would state that a monotonic

trend is contained in the dataset. The calculation of the MK test statistic, which is also known as Kendall's tau, is as follows (Yue et al., 2002):

$$S_k = \sum_{h=1}^{n-1} \sum_{i=h+1}^n \text{sign}(X_i - X_h) \tag{4}$$

$X_i$  denotes the ordered data values, and  $n$  is the length of observations; the sign test is (Yue et al., 2002):

$$\text{Sign}(X_i - X_h) = \begin{cases} +1 & \text{if } X_i > X_h \\ 0 & \text{if } X_i = X_h \\ -1 & \text{if } X_i < X_h \end{cases} \tag{5}$$

When the number of observations is greater than 10, the Kendall's tau  $S_k$  has a distribution that is approximately normal with zero mean (Hamed and Rao, 1998; Adamowski and Bougadis, 2003). The variance of the statistic  $S_k$  can then be calculated using the following equation (Kendall, 1975; Yue et al., 2002):

$$V(S_k) = \left\{ n(n-1)(2n+5) - \sum_{h=1}^{nh} t_h(h-1)(2h+5) \right\} / 18 \tag{6}$$

$t_h$  represents the number of ties or duplicates of extent  $h$  (the summation in Eq. (6) is used in the presence of tied values in the time series).  $nh$  is the total number of ties in the dataset. When  $n \geq 10$  the standardized test statistic for the Mann–Kendall test can then be calculated using (Yue et al., 2002):

$$Z = \begin{cases} \frac{S_k - 1}{\sqrt{V(S_k)}}, & \text{(if } S_k > 0) \\ 0, & \text{(if } S_k = 0) \\ \frac{S_k + 1}{\sqrt{V(S_k)}}, & \text{(if } S_k < 0) \end{cases} \tag{7}$$

Positive and negative  $Z$  values indicate that the direction of the trend is upward and downward, respectively. The calculated  $Z$  value is compared to the standard normal variate at some level of statistical significance ( $\alpha$ ) (Hamed and Rao, 1998). In a two-sided test, if the calculated  $|Z|$  is greater than  $Z_{\alpha/2}$ , it implies that there is a significant trend (i.e. the null hypothesis is rejected).

### 2.3. Modified Mann–Kendall (MK) trend tests

It is important to check for the presence of serial correlation in the time series being analyzed prior to using the original MK test. It is widely recognized that the original MK trend test should only be applied to test for a trend in a dataset that does not exhibit serial correlation (Hamed and Rao, 1998; Adamowski and Bougadis, 2003; Mohsin and Gough, 2010). If the original MK test is used on a time series that exhibits positive serial correlation, the likelihood of finding trends is enhanced, when in fact, there is no trend; and vice versa (Hirsch and Slack, 1984; Hamed and Rao, 1998). Hamed and Rao (1998) tested a time series with an AR(1) of 0.4 using the original MK test and they were able to demonstrate that the significant positive trend found (at the 5% significance level) was merely due to the effect of autocorrelation in the data.



Modifications to the original MK test have been proposed by Hirsch and Slack (1984) and Hamed and Rao (1998) in order to account for seasonality patterns and autocorrelation factors in the analyzed data, respectively. These modified versions of the MK tests are summarized in the following sections.

### 2.3.1. Modified Mann–Kendall (MK) test by Hirsch and Slack (1984) for data with seasonality patterns and autocorrelation

For a dataset,  $x$ , recorded over  $v$  season and for  $u$  years, with no missing or tied values, its matrix can be written as (Hirsch and Slack, 1984):

$$x = \begin{pmatrix} x_{11} & x_{12} & x_{13} \cdots & x_{1v} \\ x_{21} & x_{22} & x_{23} \cdots & x_{2v} \\ x_{31} & x_{32} & x_{33} \cdots & x_{3v} \\ \vdots & \vdots & \vdots & \vdots \\ x_{u1} & x_{u2} & x_{u3} \cdots & x_{uv} \end{pmatrix}. \quad (8)$$

The ranks of the data in matrix  $x$  are represented by matrix  $r$  (Hirsch and Slack, 1984):

$$r = \begin{pmatrix} r_{11} & r_{12} & r_{13} \cdots & r_{1v} \\ r_{21} & r_{22} & r_{23} \cdots & r_{2v} \\ r_{31} & r_{32} & r_{33} \cdots & r_{3v} \\ \vdots & \vdots & \vdots & \vdots \\ r_{u1} & r_{u2} & r_{u3} \cdots & r_{uv} \end{pmatrix}. \quad (9)$$

The seasonal Kendall test statistic, which is asymptotically normal with a mean of zero, is calculated using the sum of the Mann–Kendall test statistic for each season. The variance of the seasonal Kendall test statistic is then calculated by adding the sum of the variance from each season with the estimate covariance of two seasons (which was developed by Dietz and Killeen (1981)). The estimate of covariance of two seasons relies on the Spearman's correlation coefficient for cases that have no ties and no missing values. Hirsch and Slack (1984) demonstrated that by using these consistent estimators for the covariance in order to calculate the variance of the seasonal Kendall test statistic, the assumption of independence in a time series is no longer required. The details of this modified version of the MK test can be found in Hirsch and Slack (1984).

### 2.3.2. Modified Mann–Kendall (MK) test by Hamed and Rao (1998) for significantly autocorrelated data

The Hamed and Rao (1998) modified version of the MK test is intended to address the serial correlation structures in a dataset by looking at their effects on the mean and variance of the original Mann–Kendall test. An empirical approximation for the variance for the MK test was developed, which is considered suitable for autocorrelated data. The calculation of the variance of the MK test statistic is altered by incorporating the effective number of samples required to account for the autocorrelation in the dataset in to the calculation — the autocorrelation between ranks is used instead of between the data values. The details of this modified version of the MK test can be found in Hamed and Rao (1998).

Hamed and Rao (1998) applied their proposed modified version of the MK test on precipitation and streamflow series exhibiting autocorrelation and found that the power of the test is similar when compared to that of the original MK test (for independent data). The accuracy of this modified MK test is much higher than the original MK test for data that exhibit significant autocorrelation. Hamed and Rao (1998) also showed that the empirical significance level is much closer to the nominal significance level when the modified MK test is used in autocorrelated data.

## 3. Data and study sites

Data from a total of five meteorological stations located in southern Ontario and Quebec were used in this study. Harrow, Vineland, Belleville and Peterborough stations are located in Ontario and Val d'Or station is located in Quebec. For the purpose of analyzing trends associated with climate change, Kahya and Kalayci (2004) and Burn and Hag Elnur (2002) recommended that at least 31 and 25 years worth of data, respectively, are used in order to obtain a valid mean statistic. Partal (2010) considered 40 years' worth of data adequate for trend analysis studies. Additionally, Mishra and Singh (2010) considered that up to 3% of missing records are acceptable for trend analysis in meteorological studies. In light of this, the stations selected in this study have at least 40 years of data without any missing values. A number of studies looking at trends in hydroclimatic indices have also used 40 years worth of data or less (e.g. Domroes and El-Tantawi, 2005; Chaouche et al., 2010; Makokha and Shisanya, 2010; Karaburun et al., 2011).

The locations of the stations used in the study are shown in Fig. 1; and the key features of the stations are given in Table 1 — joint stations indicate that records from nearby stations were combined in order to produce longer time series. The details of how data from nearby stations were combined can be found in Mekis and Vincent (2011).

The data used in this study came from the second generation homogenized temperature data of Environment Canada. These homogenized temperature data are specially developed for trend studies in climatic indices. Adjustment procedures on monthly and daily maximum and daily minimum temperature indices were implemented to create the first generation homogenized temperature data. The adjustments were applied in order to account for non-climatic shifts such as station relocations, changes in recording procedures and automation (Vincent and Gullett, 1999). These non-climatic shifts may cause inhomogeneities in the temperature data, which in turn leads to inaccurate trend estimates if the data were to be used for trend analysis (Zhang et al., 2000). In the second generation homogenized sets, the spatial and temporal coverage of temperature data have been improved. Furthermore, additional adjustment procedures were implemented in order to solve the bias caused by the redefinition of the end time of the climatological day, which occurred as of July 1, 1961 (refer to Vincent et al., 2009 for the details of the adjustments). The adjustment procedures in the second generation homogenized datasets involved adjusting the daily minimum temperatures, which are based on hourly data for the period 1967–2006 (40 years) — the adjustment in any one day was between 0.58 and 12.58 °C (Vincent et al., 2009).



Fig. 1. The map of the weather stations used in this study.

There are four categories of temperature data analyzed in our study: monthly, seasonally-based, seasonal (i.e. winter, spring, summer, and autumn), and annual. The data spanned from 1967 to 2006, with the exception of station Harrow, whose annual and autumn data end in 2005 due to missing observations at the end of 2006.

The monthly data contain observations starting from January 1967 to December 2006 (with the exception of station Harrow whose monthly time series ends in August 2006). Seasonally-based data use the average value from each season every year continuously from winter 1967 to autumn 2006 (with the exception of station Harrow whose seasonally-based data ends in summer 2006): December–February (winter), March–May (spring), June–August (summer), and September–November (autumn). In the seasonal data analysis, each season (i.e. winter, spring, summer, and autumn) was analyzed separately.

Monthly data were analyzed in order to investigate the effects of shorter time scales (e.g. intra-annual and inter-annual periodicities) on the observed temperature trends. Seasonally-based time series were analyzed in order to investigate the effects of the semi-annual and annual seasonality on the temperature trends. Annual and seasonal time series were included in the study in order to investigate events occurring in

longer time scales such as multi-year and decadal events. Additionally, seasonal data were included because several studies found that changes in temperature do not only occur in annual data but also within the different seasons (examples for Canada: Zhang et al., 2000; Vincent et al., 2007). Karaburun et al. (2011) also indicated that an overall positive trend in the mean annual temperature data may not show that trends in some seasons may actually be negative. Therefore, it is important to analyze individual seasons separately.

#### 4. Methodology

The wavelet decomposition was applied to each time series in order to separate their high- and low-frequency components. After the decomposition, the MK trend test was then applied to the different detail components (D), approximation components (A), as well as to the detail components with their respective approximation added. The data analysis in this study was carried out using several procedures, which are summarized as follows:

1. The presence (or lack thereof) of serial correlation was checked for each dataset.
2. The presence (or lack thereof) of seasonality patterns in each dataset was determined using their correlograms.
3. Each time series was decomposed via the DWT into its details (D) and approximation (A) components. The type of mother wavelet, number of decomposition levels, and the type border extension used were determined using the relative error criterion between the MK Z-values of the original data and the approximation component at the last decomposition level (see Section 4.2).
4. The MK trend test and the sequential MK analysis were applied to the original datasets and to the different detail and approximation series produced by the wavelet decomposition.

Table 1

Key features of the meteorological stations used in this study.

Station name	Province	Station location		Elevation (m)	Joint station
		Latitude (°)	Longitude (°)		
Harrow	ON	42.0	−82.9	182	Yes
Vineland	ON	43.2	−79.4	79	Yes
Belleville	ON	44.2	−77.4	76	No
Peterborough	ON	44.2	−78.4	191	Yes
Val d'Or	QC	48.1	−77.8	337	No

5. The most important periodicities that affect the observed trends were determined by examining the sequential MK graphs and the MK Z-values of the detail (plus approximation) components, and then comparing them to that of the original data.

4.1. Serial correlation and seasonality analyses

The serial correlation test was applied in order to check if a time series exhibited non-random characteristics. If serial correlation exists in a time series, it increases the likelihood to reject the null hypothesis of no trend, when in fact the null hypothesis should be accepted (Yue et al., 2002). This is because the variance of the MK test statistic is underestimated (Hamed and Rao, 1998).

In this study, each time series' correlograms and autocorrelation coefficients (ACFs) at lag-1 were used to determine the presence (or lack thereof) of a significant autocorrelation. Lag-1 ACF is commonly used to determine whether a time series exhibit non-random characteristics (e.g. Partal and Kahya, 2006; Mohsin and Gough, 2010). Lag-1 ACFs were computed using (Yue et al., 2002; Mohsin and Gough, 2010):

$$R = \frac{\left(\frac{1}{n-1}\right) \sum_{t=1}^{n-1} [x_t - \bar{x}_t] [x_{t+1} - \bar{x}_t]}{\left(\frac{1}{n}\right) \sum_{t=1}^n [x_t - \bar{x}_t]^2} \tag{10}$$

$$\frac{\{-1 - 1.645 \sqrt{n-2}\}}{n-1} \leq R \leq \frac{\{-1 + 1.645 \sqrt{n-2}\}}{n-1} \tag{11}$$

R represents the autocorrelation coefficient at lag-1 of the time series  $x_t$ ,  $\bar{x}_t$  represents the mean of the data. If the calculated lag-1 ACF is found to be within the interval defined by Eq. (11), it can be concluded that the time series does not exhibit a significant autocorrelation. On the contrary, if the calculated lag-1 ACF is outside of the interval, it can be said

that the time series exhibits a significant autocorrelation at the 5% significance level.

The correlograms depicting the ACFs of the analyzed time series at different lags were obtained using IBM SPSS Statistics 19. If an ACF value crosses the upper or lower confidence limits, it indicates that the autocorrelation at that specific lag is significant (an example is given in Fig. 2). The correlograms were also used to identify whether a particular time series exhibits some form of seasonality. If repeated oscillating patterns that continued for many lags were observed in a correlogram, it indicates that the analyzed time series exhibited seasonality patterns (as exemplified in Figs. 2 and 3).

4.2. Discrete wavelet transform (DWT) applications on different temperature time series

Time series decomposition via the DWT was computed using the *multilevel one-dimensional wavelet analysis* function in MATLAB. The signal (i.e. time series) is convolved with low-pass and high-pass filters, followed by a dyadic discretization or downsampling procedure, in order to produce the approximation (A) and detail (D) coefficients. The signal is then reconstructed using the *multilevel one-dimensional wavelet reconstruction* function using the same band-pass filters. The original signal is decomposed into scales by powers of 2: the signal is broken down in halves, then in quarters, and it continues onward (Dong et al., 2008). The first decomposition results in detail (D1) and approximation (A1), the next iteration decomposes A1 into detail (D2) and approximation (A2); the process repeats until the desired number of decomposition levels is reached. The lower the level of the detail (D) component, the higher the frequency of information it represents. The lowest frequency information of the data is contained in the approximation (A) component of the last decomposition level. A perfect reconstruction of the original signal can then be achieved by working the calculation

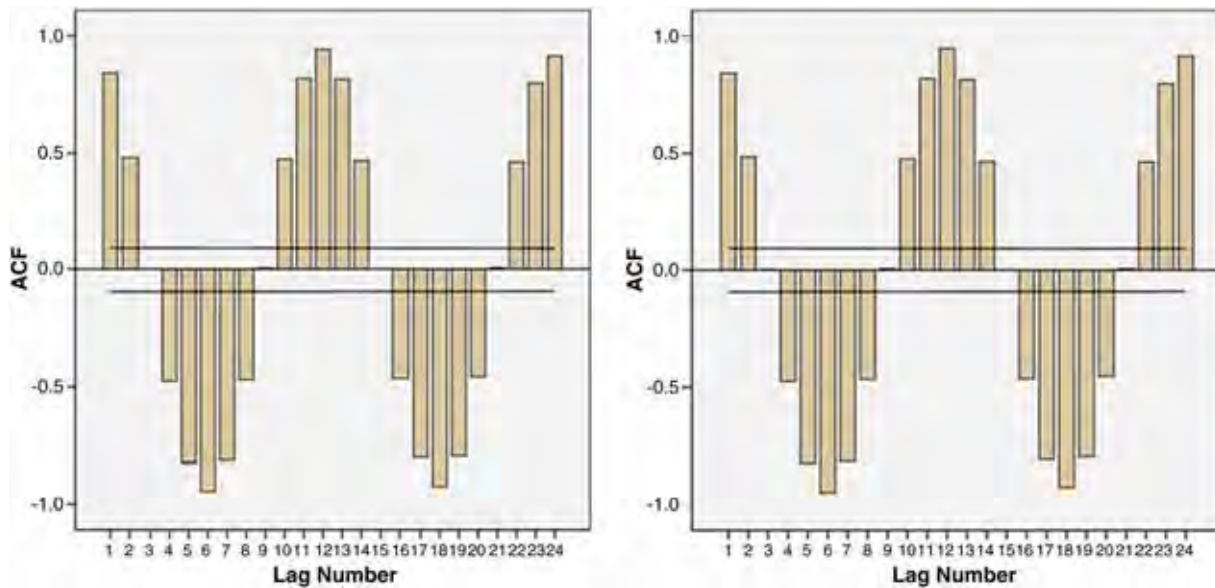


Fig. 2. Examples of the monthly data correlograms: stations Harrow (left) and Vineland (right). High coefficient values at every sixth lag indicate the presence of semi-annual and annual seasonality patterns. The upper and lower confidence limits are shown by the straight lines.

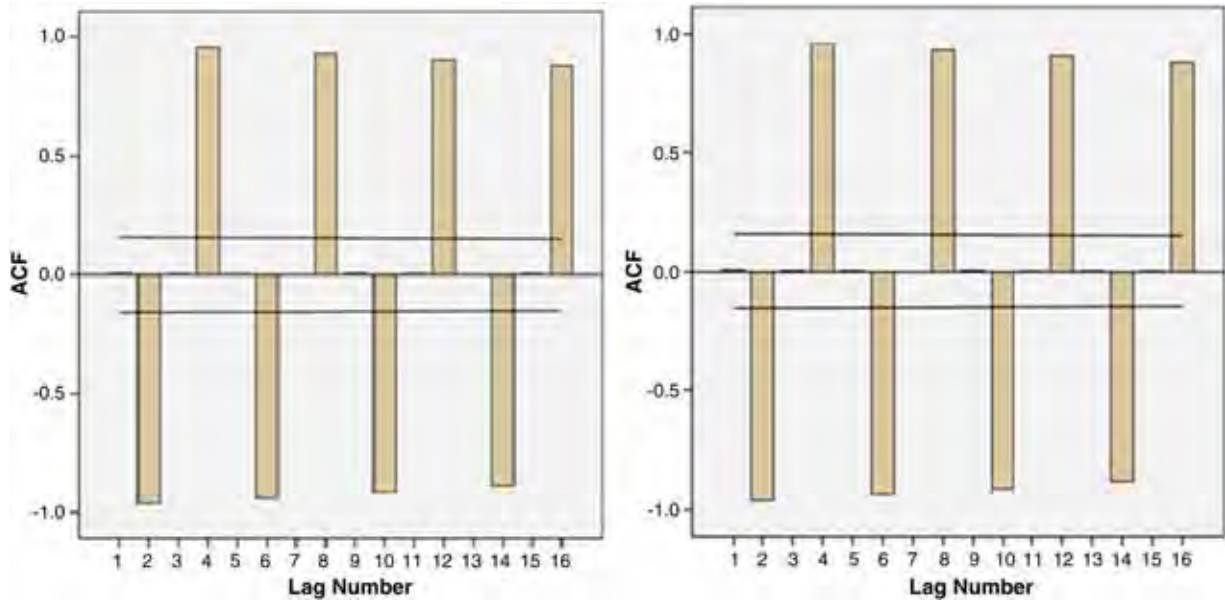


Fig. 3. Examples of the seasonally-based data correlograms: stations Harrow (left) and Vineland (right). High coefficient values at every second lag indicate the presence of semi-annual and annual seasonality patterns. The upper and lower confidence limits are shown by the straight lines.

upwards from the approximation (A) component of the last decomposition level.

Daubechies (db) wavelets were used as the mother wavelet in the time series decomposition. The Daubechies (db) wavelets were used in this study because of their ease of use, compact support, and orthogonality (Ma et al., 2003; Vonesch et al., 2007), which implies that the wavelets have non-zero basis functions over a finite interval, and also full scaling and translational orthonormality properties (Popivanov and Miller, 2002; de Artigas et al., 2006). These properties are very important for localizing events when analyzing signals that are characterized by time dependency – this localizing property also implies that wavelets can be adjusted to accommodate both high and low frequencies of the analyzed signals (Wang et al., 1998; Popivanov and Miller, 2002). Furthermore, the compact support provided by Daubechies (db) wavelets has fewer degrees of freedom (associated with the wavelet coefficients), which is ideal for analyzing signals with complex structures (Ma et al., 2003). The scaling function of a Daubechies (db) wavelet also effectively represents polynomials with order up to  $\phi/2-1$ , where  $\phi$  is an even integer (Ma et al., 2003). In order to determine the type of Daubechies (db) wavelet to be used in the time-series decomposition, this present study tried out different wavelets from db1 to db10.

Border extensions were also considered important because when performing the DWT decomposition on signals with finite length, the issue of border distortion effects is introduced. This happens because convolution processes cannot occur outside the ends of signals having limited length as there is no available information outside the ends (Su et al., 2011). Extending the ends of the signal produces several extra coefficients during the decomposition process, which are needed to ensure a perfect signal reconstruction. There are three types of border extensions that are normally used in the DWT: zero padding, periodic extension, and boundary value replication (symmetrization). Zero padding uses zeros outside of the original support of the

mother wavelet, to pad the signal being analyzed; periodic padding recovers the signal beyond the original support by periodic extension; and symmetrization – which is the default mode in MATLAB – assumes that signals outside the original support can be recovered by symmetric boundary replication (de Artigas et al., 2006). The inverse discrete wavelet transform (IDWT) was then run in MATLAB to ensure perfect signal reconstructions.

In this study, two criteria were tried out in order to calculate the number of decomposition levels and to determine the border extension and the type of Daubechies (db) mother wavelet to be used in the DWT procedure. The first criterion involved the use of the mean relative error (MRE) between the approximation (A) time series and original time series. The second criterion used the relative error (RE) between the MK Z-values of the approximation (A) of the last decomposition level and the original data. The lowest MRE and RE values were sought. The MRE was calculated using (Popivanov and Miller, 2002; de Artigas et al., 2006):

$$MRE = 1/n \sum_{j=1}^n \frac{|a_j - x_j|}{|x_j|} \quad (12)$$

where  $x_j$  is the original data value of a signal whose number of records is  $n$ , and  $a_j$  is the approximation value of  $x_j$ . The RE criterion was proposed by Nalley et al. (2012) and calculated using:

$$RE = \frac{|Z_{ap} - Z_{or}|}{|Z_{or}|} \quad (13)$$

$Z_{or}$  represents the MK Z-value of the original time series; and  $Z_{ap}$  is the MK Z-value of the approximation component of the last decomposition level of the DWT.



The following equation was proposed by de Artigas et al. (2006), who conducted a study on monthly geomagnetic activity indices, to calculate the number of decomposition levels:

$$L = \frac{\text{Log}\left(\frac{n}{2^v-1}\right)}{\text{Log}(2)} \quad (14)$$

where  $v$  represents the number of vanishing moments of a Daubechies (db) wavelet,  $n$  is the number of records in a monthly-based time series, and  $L$  is the maximum decomposition levels. In MATLAB, the number of vanishing moments for a Daubechies (db) wavelet is half of the length of its starting filter. For example, if one is using the db3 mother wavelet in MATLAB, it implies that the wavelet is Daubechies3, which has a 6-point filter length. It should also be noted that if the number of data points in a time series is not exactly in a dyadic format (as is the case in this study), the DWT computation in MATLAB is performed using the next dyadic arrangement. For example, in our study there were 480 data points in monthly datasets. The value of  $n$  in Eq. (14) would be represented by  $2^9 = 512$  (which is the next dyadic format from 480). Therefore, the number of  $n$  used in the DWT procedure would be 512. If, for example, the db3 wavelet is used on the monthly data, the number of decomposition levels,  $L$ , would be 6.68 (seven levels would then be used). Similarly, for the seasonally-based time series (having 160 data points), the computation of the DWT in MATLAB would use 256 ( $2^8$ ) as  $n$ . If db3 was used in Eq. (14) for the seasonally-based data, the calculated  $L$  would be 5.68 (six levels would then be used).

The use of the newly proposed *RE* criterion by Nalley et al. (2012) illustrated that using the *RE* was more precise in determining the most appropriate type of Daubechies (db) mother wavelet and border extension, and the number of decomposition levels. This provides a justification to use this criterion for the DWT computation in our study. When the *MRE* criterion was used to determine the number of decomposition levels (using different db types and border extensions), the differences in the *MRE* between different decomposition levels were not noticeable. For example, for station Vineland's annual data, the *MRE* for four decomposition levels using different Daubechies (db) wavelets ranged from: 0.06 to 0.07, 0.055 to 0.059, and 0.11 to 0.22 using periodic extension, symmetrization, and zero padding, respectively. The *MRE* for five decomposition levels for the same station ranged from: 0.06 to 0.08, 0.06 to 0.07, and 0.19 to 0.24 using periodic extension, symmetrization, and Zero padding, respectively. On the contrary, when the *RE* criterion was used, noticeable differences were observed. For example, for Vineland's annual data, the relative errors obtained from using four decomposition levels were: 0.01–1.58, 1.00–3.40, and 0.03–1.84 using periodic extension, symmetrization, and zero padding, respectively. For the same data, the relative errors obtained from using five decomposition levels were: 0.02–2.16, 0.58–7.66, and 0.04–2.60 using periodic extension, symmetrization, and zero padding, respectively. Therefore, for Vineland's annual data, four decomposition levels were used (the lowest *RE* of 0.01 was obtained from using db6 wavelet). This is an example of how the number of decomposition levels was determined on a case-by-case basis in this

study. The noticeable differences in the *RE* were not only seen for Vineland station, but also for all the other stations.

In our study, we observed that in most cases, using db wavelets or border extensions, other than the ones determined using the *RE* criterion, led to a different number of decomposition levels. Generally the number of decomposition levels was different by one (the analysis is not presented here). Even so, the most dominant periodicities (based on the MK Z-values of the detail components and their sequential MK graphs) may not be the same as those resulting from when the *RE* criterion was used. In addition to that, if data decomposition is done using Daubechies (db) wavelets and border extensions other than the ones determined using the *RE* criterion, the sequential MK graphs of the detail components (even for those whose MK Z-values are closest to that of the original data) are out of harmony compared to the sequential MK of the original data.

#### 4.3. The Mann–Kendall (MK) trend test

The MK test statistic  $S$  and the variance were calculated (see Eqs. (4) and (6), respectively) for each dataset in order to obtain the standard normal value,  $Z$  score (see Eq. (7)). In the data analysis of this study, the significant level used was  $\alpha = 5\%$  (or 95% confidence level) for a two-sided probability. The absolute value of this  $Z$ -score was then compared to the critical two-tailed  $Z$ -value (area under the normal curve) of  $\alpha/2$ . The  $Z$  values in a two-tailed test for  $\alpha = 5\%$  are  $\pm 1.96$ . If the calculated MK  $Z$ -score is outside the range of  $-1.96$  and  $+1.96$ , the trends are statistically significant. The MK test tests the null hypothesis of no trend (independent observations and ordered randomly) against the alternative hypothesis of positive or negative monotonic trends over time that are present in the dataset being analyzed (Hirsch and Slack, 1984; Mohsin and Gough, 2010; Karaburun et al., 2011).

##### 4.3.1. Applications of the original and modified versions of the Mann–Kendall (MK) trend test

The modified MK test by Hirsch and Slack (1984) was used on time series that exhibited seasonality patterns. The modified MK test by Hamed and Rao (1998) was used on time series that exhibited only significant autocorrelations at lag-1. The original MK test was applied to time series that exhibited neither significant autocorrelations at lag-1 nor seasonality patterns.

##### 4.3.2. Sequential Mann–Kendall (MK) analysis

The sequential MK test was used in order to examine the progressive trend lines in each time series from the beginning to the end of the study period. This is useful because positive and negative trends, which may or may not be significant, can be observed in the sequential MK graphs (Makokha and Shisanya, 2010). Additionally, with sequential MK analysis, we could also observe if a series of significant positive and negative trends may cancel each other out and thus, produce an MK  $Z$ -value that is not significant at the end of the study period. The sequential MK analysis in this study was also used to determine the periodic modes that are considered the most influential in affecting the temperature trends over the study area.

The sequential MK values were calculated using the appropriate MK test (i.e. the original or the modified versions) for each dataset, from the start to the end of the study period.

The sequential MK values were then graphed. In the sequential MK graph, the upper and lower lines correspond to the confidence limits of the standard normal Z values at  $\alpha = 5\%$ . The upper and lower confidence limits therefore, correspond to  $+1.96$  and  $-1.96$ , respectively. When the sequential MK value crosses either one of the confidence limit lines it indicates a significant trend at the 5%-significance level – crossing the upper line implies a significant positive trend, whereas crossing the lower line implies a significant negative trend.

It is important to recall that the standard normal Z-score can be used in the MK test only when the number of observations in a dataset is more than 10. With this in mind, the accuracy of the first 10 sequential MK values (i.e. up to year 1976) in the sequential MK graphs may be overlooked.

## 5. Results and discussions

### 5.1. Preliminary data analysis

#### 5.1.1. Serial correlation and seasonality factors

All the monthly data showed significant lag-1 autocorrelation coefficients (Table 2). For all other data categories, significant lag-1 autocorrelation coefficients were only observed for station Vineland's annual data ( $R = 0.35$ ) (Table 2). It is commonly expected that a monthly time series would have a stronger autocorrelation compared to its annual counterparts (Hirsch and Slack, 1984). The correlograms of all the monthly data also showed strong seasonality patterns as there are repeated patterns of cycles. Semi-annual and annual seasonality patterns are strongly apparent in all the monthly data as there are high coefficients at every 6th lag (Fig. 2). This is again confirmed by the correlograms of the seasonally-based data, where the autocorrelation functions are much higher at every 2nd lag (Fig. 3). The 2nd and 4th lags in the seasonally-based data correspond to 6 and 12 month cycles, respectively.

#### 5.1.2. The Mann–Kendall (MK) test on original data

Due to the presence of seasonality patterns in the monthly and seasonally-based data, the modified version of the MK test by Hirsch and Slack (1984) was used on these data sets. The original MK test was used on the seasonal and annual datasets that showed an absence of serial correlations. The modified MK version by Hamed and Rao (1998) was used on the annual data for Vineland station because it is the only dataset that exhibits a significant lag-1 autocorrelation.

As shown in Table 3, all of the trend values show positive signs, which indicate that all temperature indices analyzed in this study have positive trends. For the monthly, seasonally-

**Table 3**

Mann–Kendall Z-values of the original time series for the different temperature data types.

	Harrow	Vineland	Belleville	Peterborough	Val d'Or
Monthly data	3.25 <sup>a</sup>	3.39 <sup>a</sup>	3.33 <sup>a</sup>	2.45 <sup>a</sup>	2.80 <sup>a</sup>
Seasonally-based data	3.09 <sup>a</sup>	3.25 <sup>a</sup>	3.26 <sup>a</sup>	2.48 <sup>a</sup>	2.57 <sup>a</sup>
Annual data	2.88 <sup>a</sup>	3.15 <sup>a</sup>	3.58 <sup>a</sup>	2.49 <sup>a</sup>	2.18 <sup>a</sup>
Winter data	1.97 <sup>a</sup>	1.97 <sup>a</sup>	2.60 <sup>a</sup>	2.37 <sup>a</sup>	1.80
Spring data	1.67	2.15 <sup>a</sup>	1.59	1.09	1.03
Summer data	2.87 <sup>a</sup>	2.81 <sup>a</sup>	2.59 <sup>a</sup>	1.87	1.98 <sup>a</sup>
Autumn data	1.21	1.73	1.69	0.91	1.38

<sup>a</sup> Indicates a significant trend value at  $\alpha = 5\%$ .

based, and annual data analysis, all stations are experiencing statistically significant positive trends (at the 5%-level). For the seasonal data, most stations are experiencing significant positive trends for the winter season (except for station Val d'Or) and for the summer season (except for station Peterborough) – it should be noted however, that the MK Z-value of Val d'Or winter and Peterborough summer are  $+1.80$  and  $+1.87$ , respectively, which are just slightly below  $+1.96$ . Only station Vineland showed a significant trend value for the spring season; and there was no station with significant trend values for autumn.

#### 5.1.3. The number of decomposition levels for the different time series

The number of decomposition levels for each time series (Tables 4–10) was determined using the MK Z-value RE criterion. As explained in Section 4.2, the lowest relative error of the MK Z-value produced from using the combination of a specific db wavelet and a border extension was sought. As a result, different number of decomposition levels for the same temperature data category may be observed.

Since data decomposition was achieved using the DWT approach, the scales are arranged in a dyadic format (integer powers of two) from the lowest scale. Therefore, D1 represents the 2-unit periodic component, D2 represents the 4-unit periodic component, D3 represents the 8-unit periodic component, and so on. An example of time series decomposition via the DWT is given in Fig. 4. It should be noted that the MK Z-values discussed in the Results and discussions section are of the detail components (D) with their respective approximation component (A) added. The approximation (A) used was the approximation from the last decomposition level. Since the approximation components are representative of the low-frequency component (including trends) (Craigmile et al.,

**Table 2**

Lag-1 autocorrelation functions (ACFs) of the different temperature data types.

	Harrow	Vineland	Belleville	Peterborough	Val d'Or
Monthly data	0.84 <sup>a</sup> (S)	0.84 <sup>a</sup> (S)	0.84 <sup>a</sup> (S)	0.84 <sup>a</sup> (S)	0.84 <sup>a</sup> (S)
Seasonally-based data	0.004 (S)	0.008 (S)	0.006 (S)	0.003 (S)	0.001 (S)
Winter data	0.14	0.12	0.19	0.05	-0.10
Spring data	0.05	0.14	0.07	-0.01	-0.02
Summer data	0.03	0.03	-0.06	0.01	-0.10
Autumn data	-0.05	0.12	0.09	0.02	-0.04
Annual data	0.29	0.35 <sup>a</sup>	0.28	0.12	-0.004

(S) indicates the presence of seasonality.

<sup>a</sup> Indicates a significant trend value at  $\alpha = 5\%$ .

**Table 4**

Mann–Kendall Z-values of the monthly temperature series: original data, details components, approximations, and a set of combinations of the details and their respective approximations. The most effective periodic components for trends are indicated in bold format.

	Harrow	Vineleand	Belleville	Peterborough	Val d'Or
Original	3.25 <sup>a</sup>	3.39 <sup>a</sup>	3.33 <sup>a</sup>	2.45 <sup>a</sup>	Original: 2.80 <sup>a</sup>
D1	−0.37	−0.36	0.31	0.00	D1: 0.75
D2	1.49	0.90	0.53	0.71	D2: −0.02
D3	−0.35	−0.67	−0.48	−0.50	D3: 0.41
D4	0.03	0.43	0.46	0.52	D4: 0.04
D5	−0.38	0.59	0.63	0.68	D5: 0.16
D6	1.06	2.03 <sup>a</sup>	1.45	1.63	A5: 2.83 <sup>a</sup>
A6	3.32 <sup>a</sup>	3.55 <sup>a</sup>	3.24 <sup>a</sup>	2.43 <sup>a</sup>	D1 + A5: <b>3.20<sup>a</sup></b>
D1 + A6	<b>3.30<sup>a</sup></b>	<b>3.53<sup>a</sup></b>	<b>3.63<sup>a</sup></b>	<b>2.56<sup>a</sup></b>	D2 + A5: <b>3.27<sup>a</sup></b>
D2 + A6	<b>3.15<sup>a</sup></b>	3.91 <sup>a</sup>	<b>3.62<sup>a</sup></b>	<b>2.55<sup>a</sup></b>	D3 + A5: 2.18 <sup>a</sup>
D3 + A6	1.30	1.68	1.69	0.73	D4 + A5: 1.43
D4 + A6	1.17	1.74	1.57	1.15	D5 + A5: 2.10 <sup>a</sup>
D5 + A6	1.56	2.59 <sup>a</sup>	2.40 <sup>a</sup>	1.65	
D6 + A6	3.63 <sup>a</sup>	4.19 <sup>a</sup>	4.19 <sup>a</sup>	3.71 <sup>a</sup>	

<sup>a</sup> Indicates a significant trend value at  $\alpha = 5\%$ .

2004; Kallache et al., 2005), it makes sense to add them to their detail components prior to testing their trends.

*5.1.4. Determining the most dominant periodic components that affect temperature trends*

Since the main goal of this study was to determine the dominant periodic mode(s) for trend using the DWT approach, it is necessary to select the detail component(s) that best represent the trend in the analyzed data. The coefficients produced from the DWT decomposition are intermediate coefficients and thus, they need to be re-adjusted to the entire signal to determine the contribution of each frequency band to the original signal (Dong et al., 2008). The identity of the signal is contained in the approximation component (Partal, 2010). Therefore, prior to testing the trend of the detail

components, the approximation component should be added to them first – this study used the approximation component of the last decomposition level because it represents the lowest-frequency component of the signal. After doing so, the most dominant periodic components that affect the temperature trends over the study area were determined. Measuring the energy of the components resulting from the DWT decomposition has been used as a way to assess the contribution of certain wavebands in a dataset (examples are seen in Dong et al., 2008; Partal, 2010). In this study, the dominant periodic components for trends were determined in two steps. First, the MK Z-values of each detail component (with its approximation added) were compared to the MK Z-value of their respective original data. Second, the sequential MK values of each detail component (with its approximation added) were graphed along with the sequential MK values of the original data. The periodic component(s) considered the most dominant in affecting the temperature trends are the ones whose MK Z-values were closest to that of the original data and whose sequential MK graphs were the most harmonious with the sequential MK of the original data.

We also tested a number of combinations of detail components with approximation series (e.g. D1 + D2 + A) but the results produced were not conclusive (based on the observations of the MK Z-values and the sequential MK graphs). For example, station Harrow's spring temperature data has an MK Z-value of +1.67; based on the nearest MK Z-value and the sequential MK graphs (see Fig. 10), the D3 component (plus A5) is considered the most dominant periodicity for trend in mean spring temperature data (see Section 5.6 for more detail). When we combine different detail components (with approximation), even when the D3 component is part of the combination set, it does not always produce MK Z-values that are close to the MK Z-value of the original data nor does it produce good sequential MK graphs. For example, D1 + D3 + A5 only gives an MK Z-value of +0.62; D2 + D3 + A5 had an MK Z-value of only +0.57. However, D2 + D5 + A5 produced a relatively close MK Z-value of +1.48, which is close to the MK Z-value of the original data (+1.67), although neither D2 nor D5 was considered important for trends. Therefore, in this study we only chose to include

**Table 5**

Mann–Kendall Z-values of the seasonally-based temperature series: original data, details components, approximations, and a set of combinations of the details and their respective approximations. The most effective periodic components for trends are indicated in bold format.

Harrow		Vineleand		Belleville		Peterborough		Val d'or	
Original	3.09 <sup>a</sup>	Original	3.25 <sup>a</sup>	Original	3.26 <sup>a</sup>	Original	2.48 <sup>a</sup>	Original	2.57 <sup>a</sup>
D1	0.50	D1	−0.26	D1	0.31	D1	0.43	D1	0.17
D2	0.19	D2	0.31	D2	0.53	D2	−0.13	D2	0.10
D3	−0.29	D3	0.25	D3	−0.48	D3	0.15	D3	0.50
D4	0.08	D4	0.51	D4	0.46	D4	0.50	D4	0.53
D5	0.40	D5	−0.27	D5	0.63	A4	2.62 <sup>a</sup>	A4	2.58 <sup>a</sup>
D6	1.95	D6	2.01 <sup>a</sup>	D6	1.45	D1 + A4	2.84 <sup>a</sup>	D1 + A4	2.97 <sup>a</sup>
A6	2.69 <sup>a</sup>	A6	3.11 <sup>a</sup>	A6	3.24 <sup>a</sup>	D2 + A4	<b>2.36<sup>a</sup></b>	D2 + A4	2.96 <sup>a</sup>
D1 + A6	4.14 <sup>a</sup>	D1 + A6	4.06 <sup>a</sup>	D1 + A6	<b>3.63<sup>a</sup></b>	D3 + A4	<b>2.13<sup>a</sup></b>	D3 + A4	<b>2.46<sup>a</sup></b>
D2 + A6	<b>2.99<sup>a</sup></b>	D2 + A6	<b>3.16<sup>a</sup></b>	D2 + A6	<b>3.62<sup>a</sup></b>	D4 + A4	3.15 <sup>a</sup>	D4 + A4	<b>2.55<sup>a</sup></b>
D3 + A6	2.90 <sup>a</sup>	D3 + A6	<b>3.36<sup>a</sup></b>	D3 + A6	1.69				
D4 + A6	<b>3.14<sup>a</sup></b>	D4 + A6	3.70 <sup>a</sup>	D4 + A6	1.57				
D5 + A6	3.88 <sup>a</sup>	D5 + A6	3.02 <sup>a</sup>	D5 + A6	2.40 <sup>a</sup>				
D6 + A6	4.39 <sup>a</sup>	D6 + A6	4.97 <sup>a</sup>	D6 + A6	4.19 <sup>a</sup>				

<sup>a</sup> Indicates a significant trend value at  $\alpha = 5\%$ .

**Table 6**

Mann–Kendall Z-values of the annual temperature series: original data, details components, approximations, and a set of combinations of the details and their respective approximations. The most effective periodic components for trends are indicated in bold format.

Harrow		Vineland		Belleville		Peterborough		Val d'Or	
Original	2.88 <sup>a</sup>	Original	3.15 <sup>a</sup>	Original	3.58 <sup>a</sup>	Original	2.49 <sup>a</sup>	Original	2.18 <sup>a</sup>
D1	−0.15	D1	0.55	D1	0.36	D1	0.22	D1	−0.01
D2	0.31	D2	0.59	D2	0.48	D2	0.80	D2	1.04
D3	0.51	D3	1.25	D3	0.69	D3	0.52	D3	−0.13
D4	0.68	D4	2.92 <sup>a</sup>	D4	3.23 <sup>a</sup>	D4	0.85	D4	1.78
D5	3.70 <sup>a</sup>	A4	3.11 <sup>a</sup>	D5	3.32 <sup>a</sup>	D5	2.37 <sup>a</sup>	D5	2.50 <sup>a</sup>
A5	2.85 <sup>a</sup>	D1 + A4	2.37 <sup>a</sup>	A5	3.72 <sup>a</sup>	A5	2.37 <sup>a</sup>	A5	2.37 <sup>a</sup>
D1 + A5	1.62	D2 + A4	2.27 <sup>a</sup>	D1 + A5	1.29	D1 + A5	0.66	D1 + A5	0.17
D2 + A5	1.98 <sup>a</sup>	D3 + A4	<b>2.90<sup>a</sup></b>	D2 + A5	2.02 <sup>a</sup>	D2 + A5	2.78 <sup>a</sup>	D2 + A5	1.32
D3 + A5	<b>2.71<sup>a</sup></b>	D4 + A4	4.93 <sup>a</sup>	D3 + A5	1.92	D3 + A5	1.41	D3 + A5	0.41
D4 + A5	3.80 <sup>a</sup>			D4 + A5	4.42 <sup>a</sup>	D4 + A5	<b>2.16<sup>a</sup></b>	D4 + A5	<b>2.25<sup>a</sup></b>
D5 + A5	4.94 <sup>a</sup>			D5 + A5	<b>3.23<sup>a</sup></b>	D5 + A5	5.28 <sup>a</sup>	D5 + A5	3.88 <sup>a</sup>

<sup>a</sup> Indicates a significant trend value at  $\alpha = 5\%$ .

**Table 7**

Mann–Kendall Z-values of the winter temperature series: original data, details components, approximations, and a set of combinations of the details and their respective approximations. The most effective periodic components for trends are indicated in bold format.

Harrow		Vineland		Belleville		Peterborough		Val d'Or	
Original	1.97 <sup>a</sup>	Original	1.97 <sup>a</sup>	Original	2.60 <sup>a</sup>	Original	2.37 <sup>a</sup>	Original	1.80
D1	−0.10	D1	0.06	D1	0.24	D1	0.01	D1	0.22
D2	0.29	D2	−0.13	D2	0.41	D2	0.15	D2	−0.06
D3	−0.20	D3	0.90	D3	0.06	D3	0.38	D3	−0.38
A3	2.09 <sup>a</sup>	D4	0.41	D4	2.69 <sup>a</sup>	D4	0.83	D4	0.85
D1 + A3	<b>1.88<sup>a</sup></b>	D5	2.37 <sup>a</sup>	D5	3.95 <sup>a</sup>	D5	2.34 <sup>a</sup>	D5	2.39 <sup>a</sup>
D2 + A3	2.69 <sup>a</sup>	A5	2.37 <sup>a</sup>	A5	2.74 <sup>a</sup>	A5	2.37 <sup>a</sup>	A5	2.37 <sup>a</sup>
D3 + A3	<b>2.23<sup>a</sup></b>	D1 + A5	1.08	D1 + A5	1.08	D1 + A5	1.13	D1 + A5	0.52
		D2 + A5	1.11	D2 + A5	1.55	D2 + A5	1.34	D2 + A5	0.80
		D3 + A5	<b>1.41</b>	D3 + A5	<b>1.64</b>	D3 + A5	1.60	D3 + A5	0.71
		D4 + A5	<b>1.76</b>	D4 + A5	5.04 <sup>a</sup>	D4 + A5	<b>2.39<sup>a</sup></b>	D4 + A5	<b>2.27<sup>a</sup></b>
		D5 + A5	5.28	D5 + A5	4.91 <sup>a</sup>	D5 + A5	5.21 <sup>a</sup>	D5 + A5	5.28 <sup>a</sup>

<sup>a</sup> Indicates a significant trend value at  $\alpha = 5\%$ .

analysis on individual detail components (with their respective approximation components added).

## 5.2. Monthly temperature data analysis

As shown in Table 4, all stations are experiencing significant positive trends. The results of the MK test showed that none of the individual detail components showed significant MK Z-values, except for the D6 component of

station Vineland ( $Z = +2.03$ ) (Table 4). After the addition of the approximation components to their respective details, it is observed that most of the trend values became significant ( $\alpha = 5\%$ ). By examining the sequential MK graphs, and by comparing the MK Z-values of the detail components and the original data, it is found that the periodic components that are effective for trends are relatively similar for all stations. Fig. 5 is an example illustrating how the most dominant periodic component(s) were chosen. Although graphically as

**Table 8**

Mann–Kendall Z-values of the spring temperature series: original data, details components, approximations, and a set of combinations of the details and their respective approximations. The most effective periodic components for trends are indicated in bold format.

Harrow		Vineland		Belleville		Peterborough		Val d'Or	
Original	1.67	Original	2.15 <sup>a</sup>	Original	1.59	Original	1.09	Original	1.03
D1	0.69	D1	0.22	D1	0.83	D1	0.10	D1	0.48
D2	−0.08	D2	0.43	D2	−0.10	D2	0.08	D2	1.11
D3	0.78	D3	1.13	D3	−0.15	D3	0.41	D3	−0.85
D4	7.38 <sup>a</sup>	A3	2.19 <sup>a</sup>	D4	3.58 <sup>a</sup>	A3	1.25	D4	1.97
D5	6.47 <sup>a</sup>	D1 + A3	2.53 <sup>a</sup>	D5	0.24	D1 + A3	0.55	A4	0.92
A5	1.67	D2 + A3	<b>1.99<sup>a</sup></b>	A5	1.67	D2 + A3	0.31	D1 + A4	<b>0.52</b>
D1 + A5	0.73	D3 + A3	3.25 <sup>a</sup>	D1 + A5	<b>0.92</b>	D3 + A3	<b>0.71</b>	D2 + A4	1.88
D2 + A5	0.15			D2 + A5	0.36			D3 + A4	0.00
D3 + A5	<b>0.87</b>			D3 + A5	0.08			D4 + A4	2.69
D4 + A5	7.10 <sup>a</sup>			D4 + A5	3.97 <sup>a</sup>				
D5 + A5	6.05 <sup>a</sup>			D5 + A5	0.80				

<sup>a</sup> Indicates a significant trend value at  $\alpha = 5\%$ .



**Table 9**

Mann–Kendall Z-values of the summer temperature series: original data, details components, approximations, and a set of combinations of the details and their respective approximations. The most effective periodic components for trends are indicated in bold format.

Harrow		Vineland		Belleville		Peterborough		Val d'or	
Original	2.87 <sup>a</sup>	Original	2.81 <sup>a</sup>	Original	2.59 <sup>a</sup>	Original	1.87	Original	1.98 <sup>a</sup>
D1	0.66	D1	0.85	D1	−0.29	D1	0.78	D1	−0.50
D2	0.55	D2	0.45	D2	0.59	D2	−0.24	D2	0.38
D3	2.23 <sup>a</sup>	D3	2.53 <sup>a</sup>	D3	1.39	D3	1.69	D3	1.15
D4	2.25 <sup>a</sup>	D4	−0.38	D4	5.35 <sup>a</sup>	D4	−0.78	D4	4.53 <sup>a</sup>
D5	3.62 <sup>a</sup>	D5	3.32 <sup>a</sup>	A4	2.35 <sup>a</sup>	A4	1.90	A4	1.85
A5	2.74 <sup>a</sup>	A5	2.74 <sup>a</sup>	D1 + A4	0.48	D1 + A4	1.39	D1 + A4	−0.08
D1 + A5	1.50	D1 + A5	1.53	D2 + A4	1.67	D2 + A4	0.57	D2 + A4	0.92
D2 + A5	<b>1.92</b>	D2 + A5	1.62	D3 + A4	<b>2.25<sup>a</sup></b>	D3 + A4	2.57 <sup>a</sup>	D3 + A4	<b>2.09<sup>a</sup></b>
D3 + A5	<b>3.97<sup>a</sup></b>	D3 + A5	4.09 <sup>a</sup>	D4 + A4	5.23 <sup>a</sup>	D4 + A4	<b>1.74</b>	D4 + A4	2.88 <sup>a</sup>
D4 + A5	4.70 <sup>a</sup>	D4 + A5	<b>3.41<sup>a</sup></b>						
D5 + A5	5.00 <sup>a</sup>	D5 + A5	4.58 <sup>a</sup>						

<sup>a</sup> Indicates a significant trend value at  $\alpha = 5\%$ .

shown in Fig. 5, all detail components show harmonious trend lines, details that have the closest MK Z-values to that of the original data are D1 and D2. The 2-month and 4-month periodic components are the most dominant components for trends in the monthly temperature for stations Harrow, Belleville, Peterborough, and Val d'Or; station Vineland's most dominant component for trends is the 2-month periodicity (Table 4).

As shown in Table 4, the trends for the monthly data in all stations seem to be affected by high-frequency components ranging from 2 to 4 months. Since the data are based on daily measurements, there could be many daily (high-frequency) variations that contribute to the trends in these higher resolution data. Examples of these daily variations are: variation in solar radiation (which can be associated with seasonality), cloud cover, albedo, air moisture content, soil heat capacity, and atmospheric wind movements that can have significant effects on the diurnal temperature (Gough, 2008). Gough (2008) also emphasized that in mid-latitude regions, mid-latitude cyclones may produce temperature clusters whose effects may last for a month. These daily variations may be very strong, and thus, conceal the effects of low-frequency periodicities (i.e. the higher detail components of the DWT).

Although none of the most dominant periodic components for any of the stations are between 6 and 12 months, it is still worthwhile to investigate their seasonally-based data, in order to investigate whether the semi-annual and annual

seasonality may be contributing to the observed warming trends in temperature over the study area.

### 5.3. Seasonally-based temperature data analysis

In this section, particular attention is given to the D1 and D2 detail components because they represent the 6-month and 12-month periodicities, which are assumed to be associated with the seasonality factor observed in the monthly and seasonally-based data. As shown in Table 5, the D2 component is the most frequently observed as the most dominant periodic component affecting trends — except for Val d'Or, all stations have D2 as one of the most dominant periodic components. Fig. 6 illustrates the use of the sequential MK analysis in determining the most harmonious detail component (with the approximation added) for the seasonally-based temperature data. As can be seen in Table 5 and Fig. 6, the yearly fluctuations, which are represented by the D2 component, are contributing in affecting the warming trends in temperature over the study area.

### 5.4. Annual temperature data analysis

Analysis on Canada's annual mean temperature has shown that warming trends are apparent nation-wide. Zhang et al. (2000) reported an increase of 0.5–1.5 °C over the 20th century. Vincent et al. (2007) found that the annual mean temperature in Canada increased by 1.2 °C over the period

**Table 10**

Mann–Kendall Z-values of the autumn temperature series: original data, details components, approximations, and a set of combinations of the details and their respective approximations. The most effective periodic components for trends are indicated in bold format.

Harrow		Vineland		Belleville		Peterborough		Val d'Or	
Original	1.21	Original	1.73	Original	1.69	Original	0.91	Original	1.38
D1	0.15	D1	0.43	D1	0.03	D1	−0.03	D1	−0.03
D2	−0.05	D2	0.76	D2	0.01	D2	0.27	D2	−0.08
D3	−0.07	D3	−1.50	D3	0.48	D3	−0.15	D3	0.18
D4	0.8	A3	1.88	D4	0.66	D4	−0.27	A3	1.51
A4	1.16	D1 + A3	2.16 <sup>a</sup>	A4	1.88	A4	0.85	D1 + A3	<b>1.53</b>
D1 + A4	1.81	D2 + A3	2.06 <sup>a</sup>	D1 + A4	<b>1.48</b>	D1 + A4	1.11	D2 + A3	1.72
D2 + A4	<b>1.31</b>	D3 + A3	<b>1.76</b>	D2 + A4	1.90	D2 + A4	<b>1.01</b>	D3 + A3	1.68
D3 + A4	2.69 <sup>a</sup>			D3 + A4	1.90	D3 + A4	0.76		
D4 + A4	1.86			D4 + A4	2.46 <sup>a</sup>	D4 + A4	−0.48		

<sup>a</sup> Indicates a significant trend value at  $\alpha = 5\%$ .

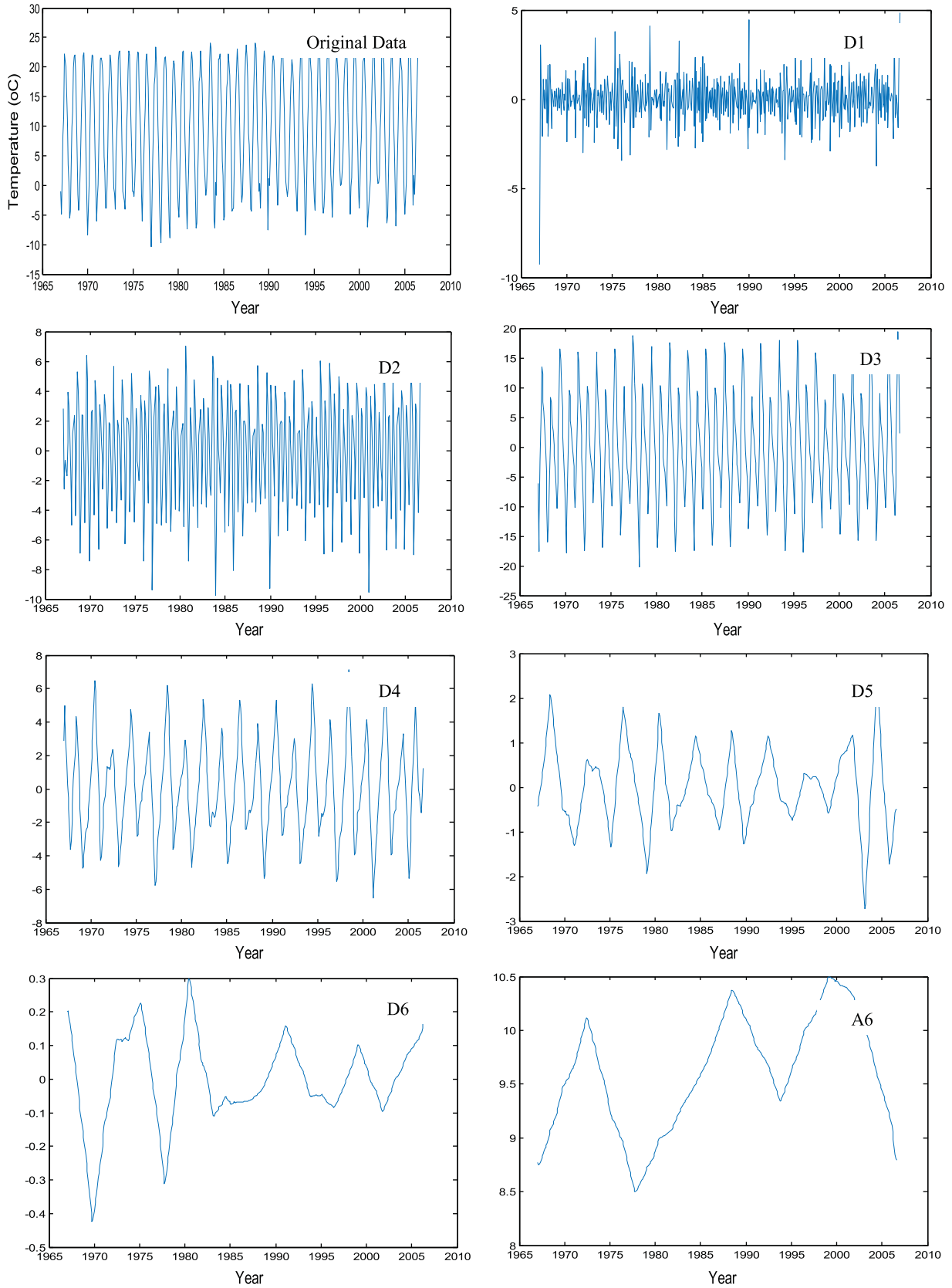
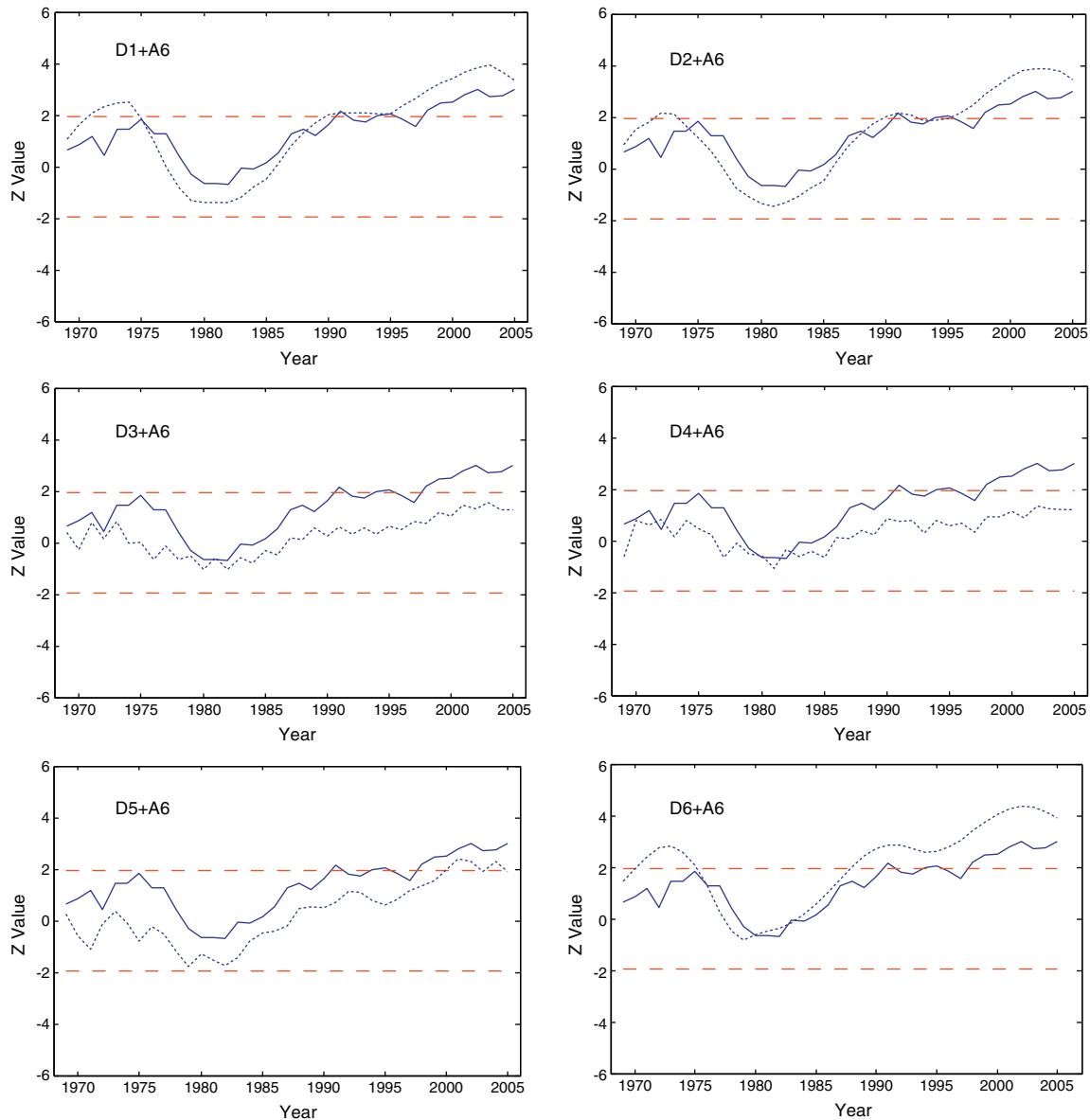


Fig. 4. Station Harrow's monthly temperature series and its decomposition via the DWT using db3 wavelet, into six levels (D1–D6 and A6).

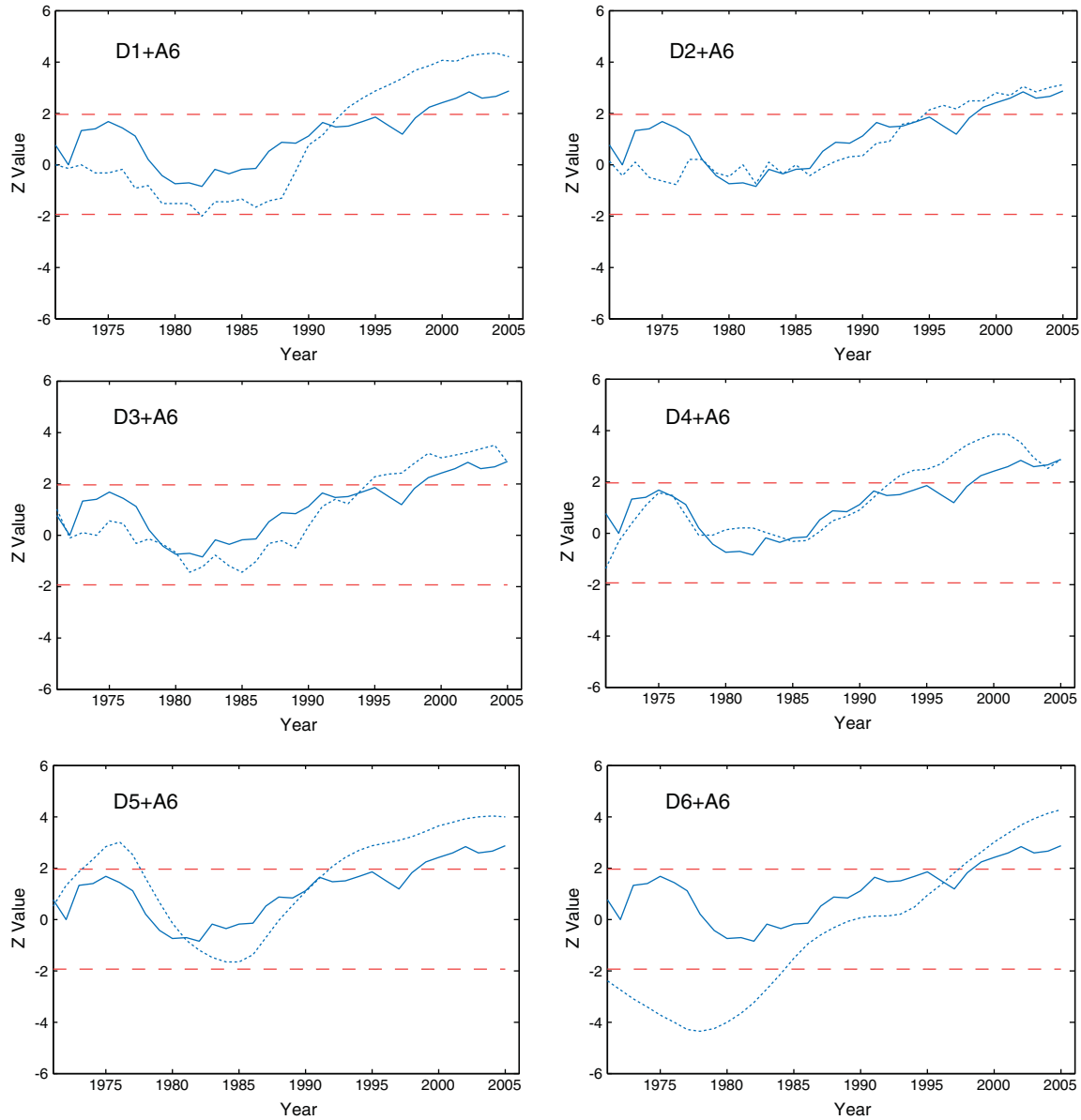


**Fig. 5.** Sequential Mann-Kendall graphs of station Harrow's monthly temperature data. The progressive trend lines of the original data are represented by the solid lines and the trend lines of the detail components (with their approximation added) are represented by the dashed line. The upper and lower dashed lines represent the confidence limits ( $\alpha = 5\%$ ). For this station, D1 and D2 are considered the most dominant periodic components because of the harmony of their progressive trend lines and their MK Z-values being close to that of the original data.

1955–2005. A more recent assessment by [Statistics Canada \(2011\)](#) also showed that over the period of 1948 to 2009 there was an increase of  $1.4\text{ }^{\circ}\text{C}$  in the mean annual temperature in Canada. More pertinent to our study, the Great Lakes and St. Lawrence and the Northeastern Forest (which covers most of Ontario and Quebec) regions experienced an increasing trend in the mean temperature departure from 1961 to 1990 normal – the mean temperature trend increased up to  $0.9\text{ }^{\circ}\text{C}$  over the period 1948–2009 ([Statistics Canada, 2011](#)).

In our study, we also discovered that all the MK Z-values of the annual data are greater than  $+1.96$ , which imply that there are significant positive trends. The annual temperature

data were decomposed into either 4 or 5 levels. For stations Harrow and Vineland, the D3 components – which correspond to the 8-year periodicity – were considered the most dominant periodic modes that affect the temperature trends in the annual data ([Table 6](#)). The MK Z-values of the D3 components for these stations are the closest to the MK Z-values of their corresponding original data. Furthermore, the sequential MK graphs of the D3 components are also harmonious with those of the original data (see [Fig. 7](#) for example). For station Belleville, the 32-year periodic mode is the most dominant one, and for stations Peterborough and Val d'Or, it is the 16-year mode. To summarize, the increasing



**Fig. 6.** Sequential Mann–Kendall graphs of station Harrow’s seasonally-based temperature data. The progressive trend lines of the original data are represented by the solid lines and the trend lines of the detail components (with their approximation added) are represented by the dashed lines.

trends in the annual mean temperature during 1954–2008 over southern parts of Quebec and Ontario are affected by periodicities between 8 and 32 years.

Since all annual temperature data in this study showed significant trend values, each season was analyzed separately in order to investigate the seasons that contribute to the warming trend over the study area to a greater extent compared to the other seasons.

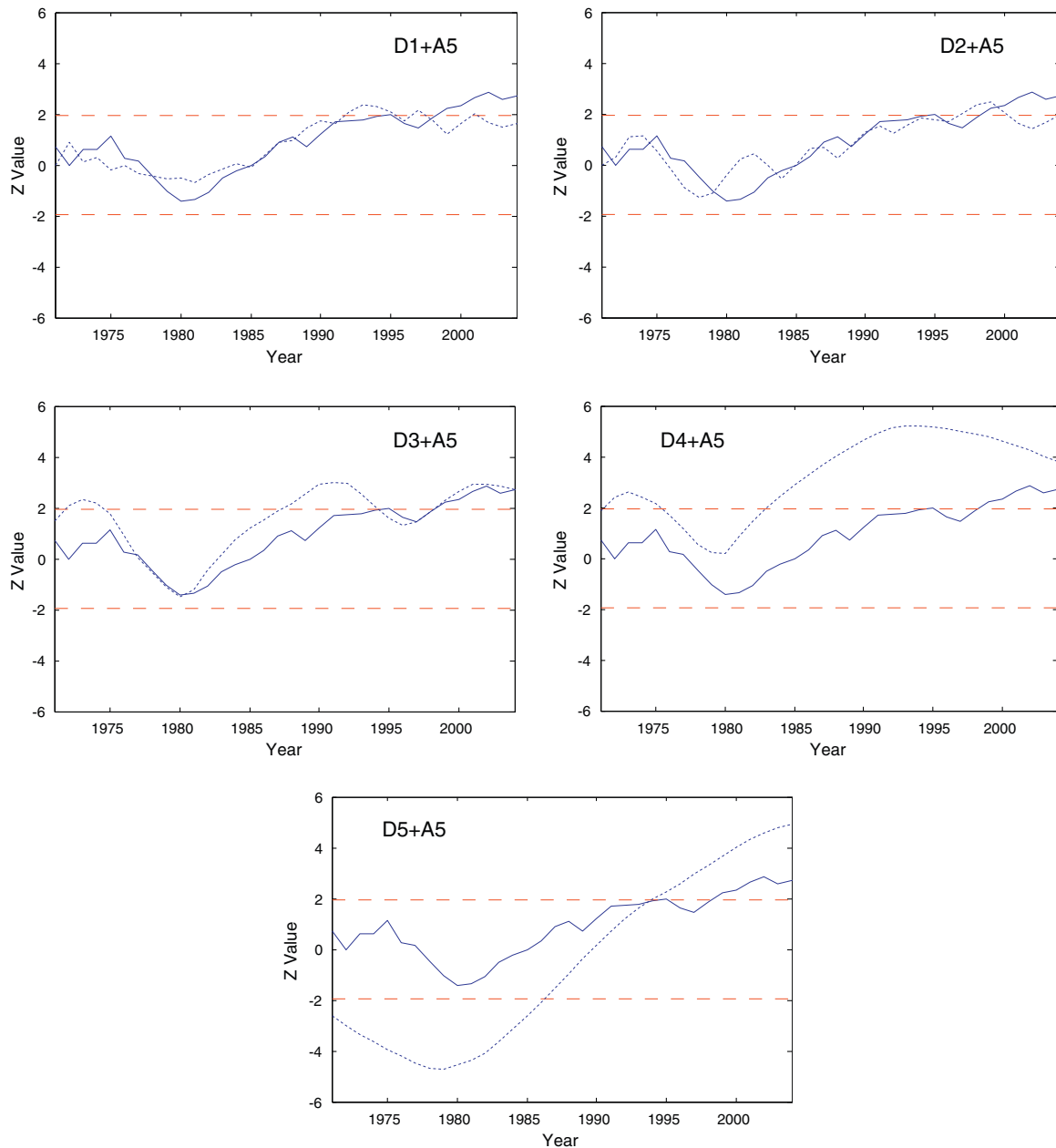
*5.5. Winter temperature data analysis*

Several studies have mentioned that winter experiences significant warming trends in the northern hemisphere and in countries such as Canada and the USA (e.g. Jones and Briffa,

1992; Lu et al., 2005; Vincent et al., 2007; Mohsin and Gough, 2010; Bukovsky, 2012). In this study, it is also observed that winter warming is very apparent because, apart from station Val d’Or, all stations show significant positive trends with MK Z-values that are relatively high. Even for Val d’Or, the winter MK Z-value (+ 1.80) is also just slightly below + 1.96.

The winter time series for station Harrow was decomposed into three levels, and the remaining time series were decomposed into five levels (Table 7). Table 7 summarizes the MK Z-values for the winter temperature data decompositions, as well as the periodic modes that are considered most dominant for winter temperature trends. The winter temperature trends for station Harrow are mostly affected by the 2-year and 8-year periodicities (i.e. D1 and D3 detail components) (Fig. 8).



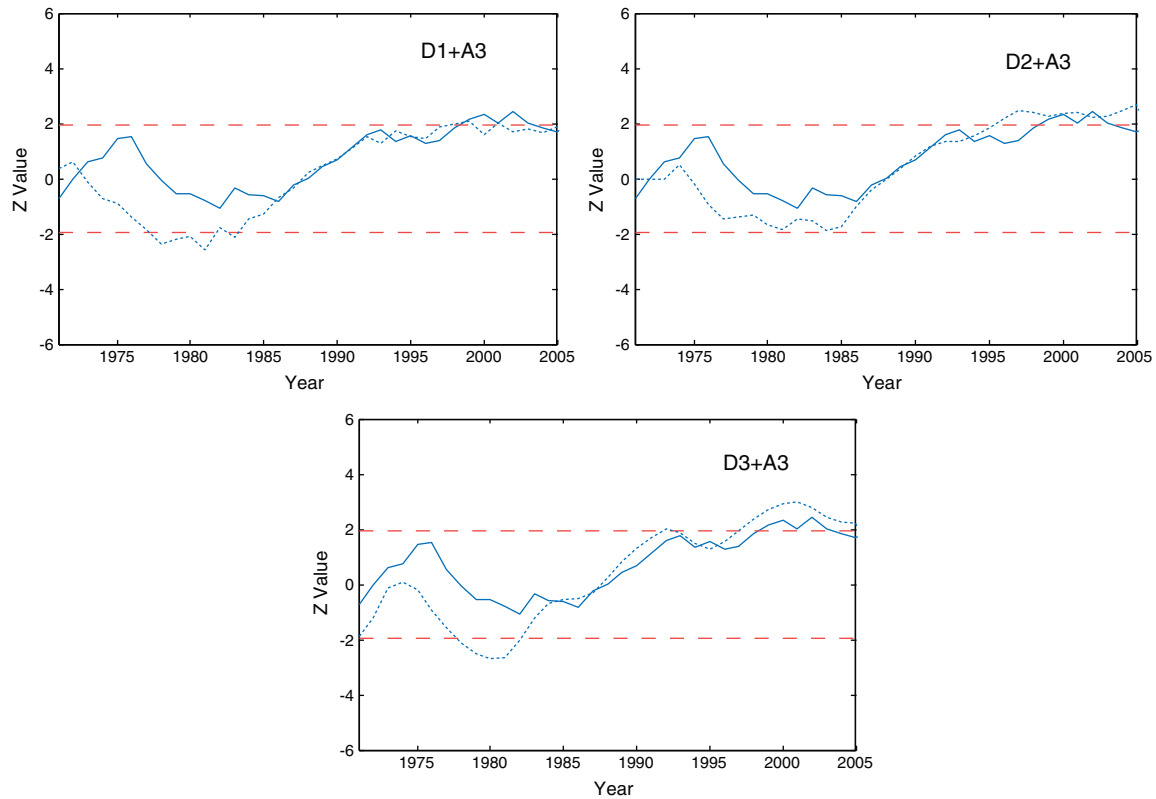


**Fig. 7.** Sequential Mann–Kendall graphs of station Harrow's annual temperature data. The progressive trend lines of the original data are represented by the solid lines and the trend lines of the detail components (with their approximation added) are represented by the dashed lines. The upper and lower dashed lines represent the confidence limits ( $\alpha = 5\%$ ). For this station, D3, which represents the 8-year time scale, is the most dominant periodicity for trends.

Peterborough station's most dominant periodicity is the 16-year periodic mode (i.e. D4 component with approximation). For station Vineland, D3 and D4 (with approximation) have MK Z-values that are closest to that of the original data ( $Z = +1.97$ ); as well, graphically, D3 and D4 have good sequential harmony with the original data compared to the rest (Fig. 9). Therefore, it can be said that for station Vineland, the 8-year and 16-year periodic components are considered the most dominant for trends in winter temperature data. Similarly, for station Belleville, the D3 component (8-year

periodic mode) has the closest MK Z-value ( $Z = +1.64$ ) to that of the original data ( $Z = +2.60$ ) with better sequential MK compared to other detail components.

Similar to the results of the annual data analysis, the trends in winter temperature warming are also mostly affected by periodic events of 8 years or greater (up to 16 years). These important periodicities may be related to the variability of the large-scale atmospheric circulations such as the North Atlantic Oscillation (NAO), El-Niño Southern Oscillation (ENSO), and Pacific North American (PNA) oscillation. The NAO is a very



**Fig. 8.** Sequential Mann–Kendall graphs of station Harrow's winter temperature data. The progressive trend lines of the original data are represented by the solid lines and the trend lines of the detail components (with their approximation added) are represented by the dashed lines. The upper and lower dashed lines represent the confidence limits ( $\alpha = 5\%$ ). For this station, D1 and D3 components were considered the most dominant periodicities for trends.

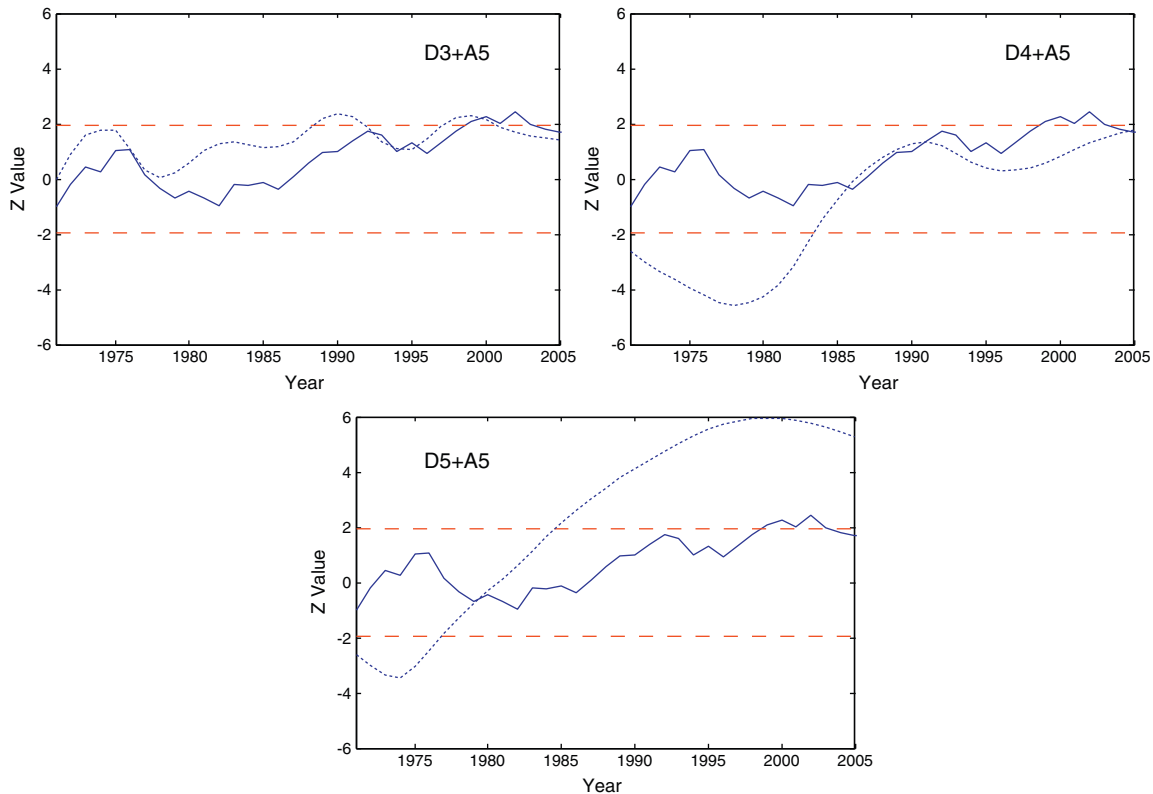
important large-scale climatic phenomenon in the northern hemisphere (Anctil and Coulibaly, 2004), especially in central and eastern Canada (Damyanov et al., 2012) and is known to strongly exhibit inter-annual to decadal variability with some of its major peaks centered around 2.1, 8 and 24 years (Cook et al., 1998; Anctil and Coulibaly, 2004). Many studies have also found a strong relationship between the NAO and temperature trends (e.g. for Canada: Wettstein and Mearns, 2002; Bonsal et al., 2006; Damyanov et al., 2012). Positive phases of the NAO cycles tend to cause above normal temperatures. The NAO has been in its positive phase since 1970 (Anctil and Coulibaly, 2004), which could contribute to the significant positive trends observed in this study. Hasanean (2001) also mentioned that the NAO variability is strongest during winter, and that the winter NAO cycle is very effective in affecting temperature variability in mid-latitude areas. Other important large-scale climate circulations such as the ENSO and the PNA oscillation also affect the temperature trends (e.g. Bonsal and Shabbar, 2011). One of the causes of the variability of these large-scale climate circulations is related to solar activities, which are frequently manifested as the 11-year solar period (Prokoph et al., 2012). It has been indicated that there are similar variabilities between surface temperatures and the 11-year solar cycle, which may contribute to the observed global warming to some extent (e.g. Lassen, 1991; Erykin et al., 2009; de Jager et al., 2010; Solheim et al., 2011). The 11-year solar period could also be applicable in our study since the periodicity

is in between 8- and 16-year modes, which are the most commonly observed as the dominant periodicities affecting the temperature trends.

#### 5.6. Spring temperature data analysis

All stations showed positive MK Z-values with only Vineland experiencing a significant trend. The number of decomposition levels via the DWT for each spring temperature data can be seen in Table 8. As shown in Table 8, the D3 (8-year periodicity) component is the one considered most influential for the spring temperature trend in stations Harrow and Peterborough. For stations Vineland, it is the D2 (4-year periodicity), and for stations Belleville and Val d'Or, the D1 (2-year periodicity) is the most dominant for trend.

An example of spring temperature analysis using the DWT and sequential MK analysis is shown in Fig. 10. It is noted that for station Harrow, the dominant 8-year periodicity is consistent with the observations in the winter and annual data. Even though most of the trend values for spring temperature are not significant, all of the trend values are positive. It is also possible that positive and negative trends may cancel each other out at some point over the study period. Determining the most dominant periodic modes for trends is still deemed important because it helps to understand the periodicities that characterize the trends in spring temperature.



**Fig. 9.** A comparison of the sequential Mann–Kendall graphs among D3, D4, and D5 (all with approximation added) of station Vineland's winter temperature data. The progressive trend lines of the original data are represented by the solid lines and the trend lines of the detail components are represented by the dashed lines. The upper and lower dashed lines represent the confidence limits ( $\alpha = 5\%$ ).

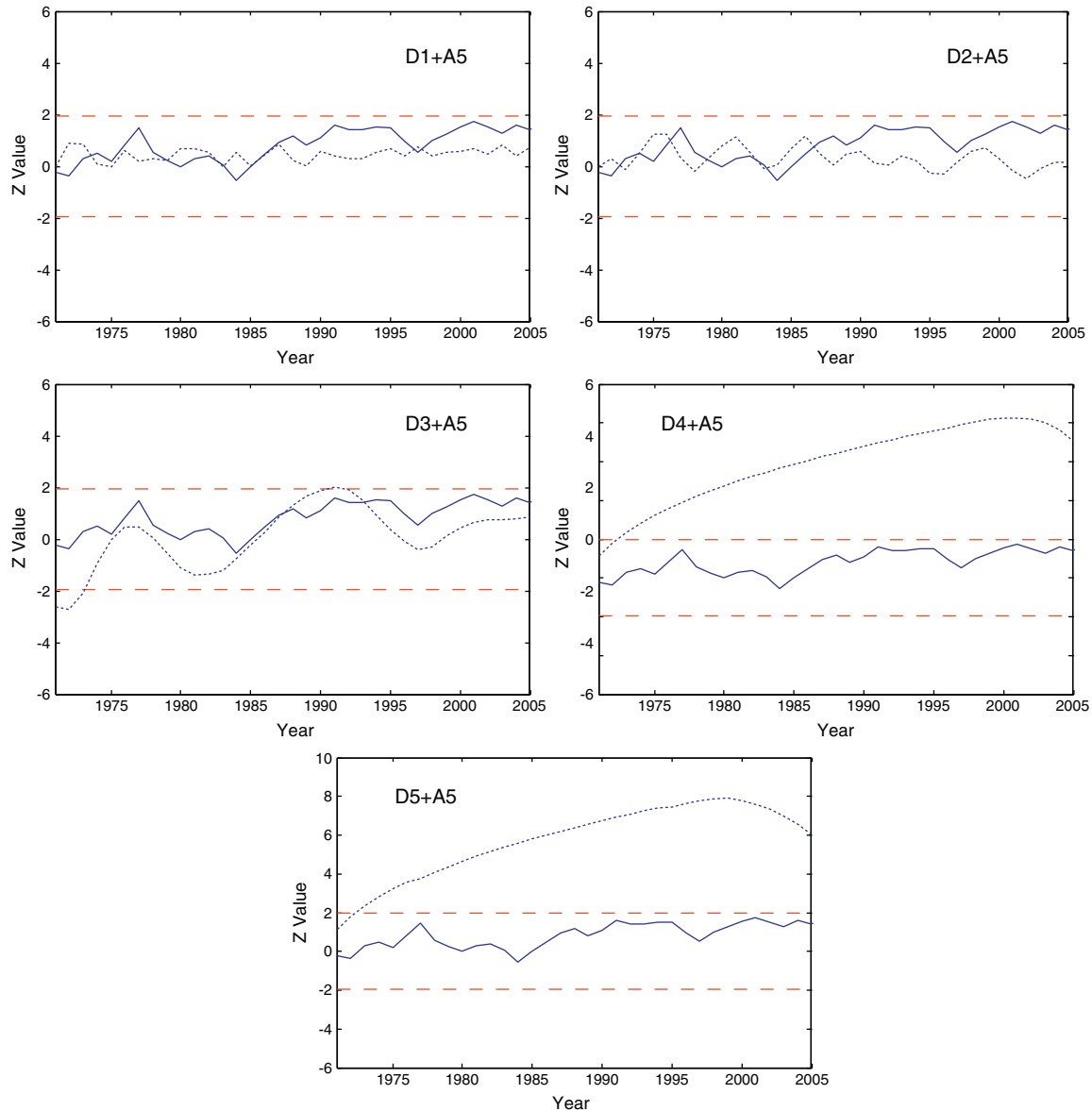
The finding in our study showing that most stations did not experience a significant temperature increase for the spring season is somewhat inconsistent with the findings from several studies that analyzed seasonal temperatures in mid-latitude areas (including Canada). Zhang et al. (2000) found that spring experienced the greatest warming in southern Canada. Vincent et al. (2007) also emphasized a significant warming in the spring in southern Canada during the period 1953–2005. This disagreement could be caused by the differences in the geographical locations of the stations used and the time period chosen. The stations included in our study are only concentrated around the most south-westerly parts of Ontario and Quebec. Even so, these differences suggest that it is important to conduct a more localized assessment on trends in temperature.

### 5.7. Summer temperature data analysis

In addition to winter, it has also been pointed out by other Canadian studies that summer also experiences significant warming, although sometimes to a lesser extent compared to winter warming (e.g. Vincent et al., 2007; Mohsin and Gough, 2010). In this study, all stations are experiencing significant positive summer temperature trends, except for Peterborough station (Table 9). Table 9 summarizes the decomposition of the summer temperature time series and the MK Z-values of the different detail components, and the details plus their respective approximations. As can be seen in Table 9, the most dominant

periodicities for trends are the D3 and D4 components, which represent the 8-year and 16-year time periodicities, respectively.

Fig. 11 shows an example of the sequential MK analysis on Harrow's summer temperature series, and how the D2 and D3 components (with approximation) show the most harmonious trend lines as compared to that of the original data. It is interesting to note that the most dominant periodicities in stations Harrow and Peterborough are again consistent with the results from their annual and winter data analysis: D3 for Harrow and D4 for Peterborough. It is again seen that in the summer data analysis, the most influential periodic modes that affect the trends are made up of multi-year and decadal events (between 8 and 16 years). The agreement found in annual, winter, and summer temperature trends could suggest that the positive trends in the annual temperature over the study area may be contributed mostly by the increase in winter and summer temperatures. Several causes of winter and summer warming trends in Canada have been investigated in past studies. For example, Vincent et al. (2007) concluded that winter warming in Canada is due to an increase in dewpoint and specific humidity. More specifically, Prokoph and Patterson (2004) and Adamowski and Prokoph (2013) associated winter warming in urban settings in eastern Ontario with the heat island effect. Summer warming is also associated with the increase in air moisture, especially around the Great Lakes and St. Lawrence areas (Vincent et al., 2007).



**Fig. 10.** Sequential Mann–Kendall graphs of station Harrow's spring temperature data. The progressive trend lines of the original data are represented by the solid line and the trend lines of the detail components (with their approximation added) are represented by the dashed lines. The upper and lower dashed lines represent the confidence limits ( $\alpha = 5\%$ ). For this station, the D3 component was considered the most dominant periodicity for trends.

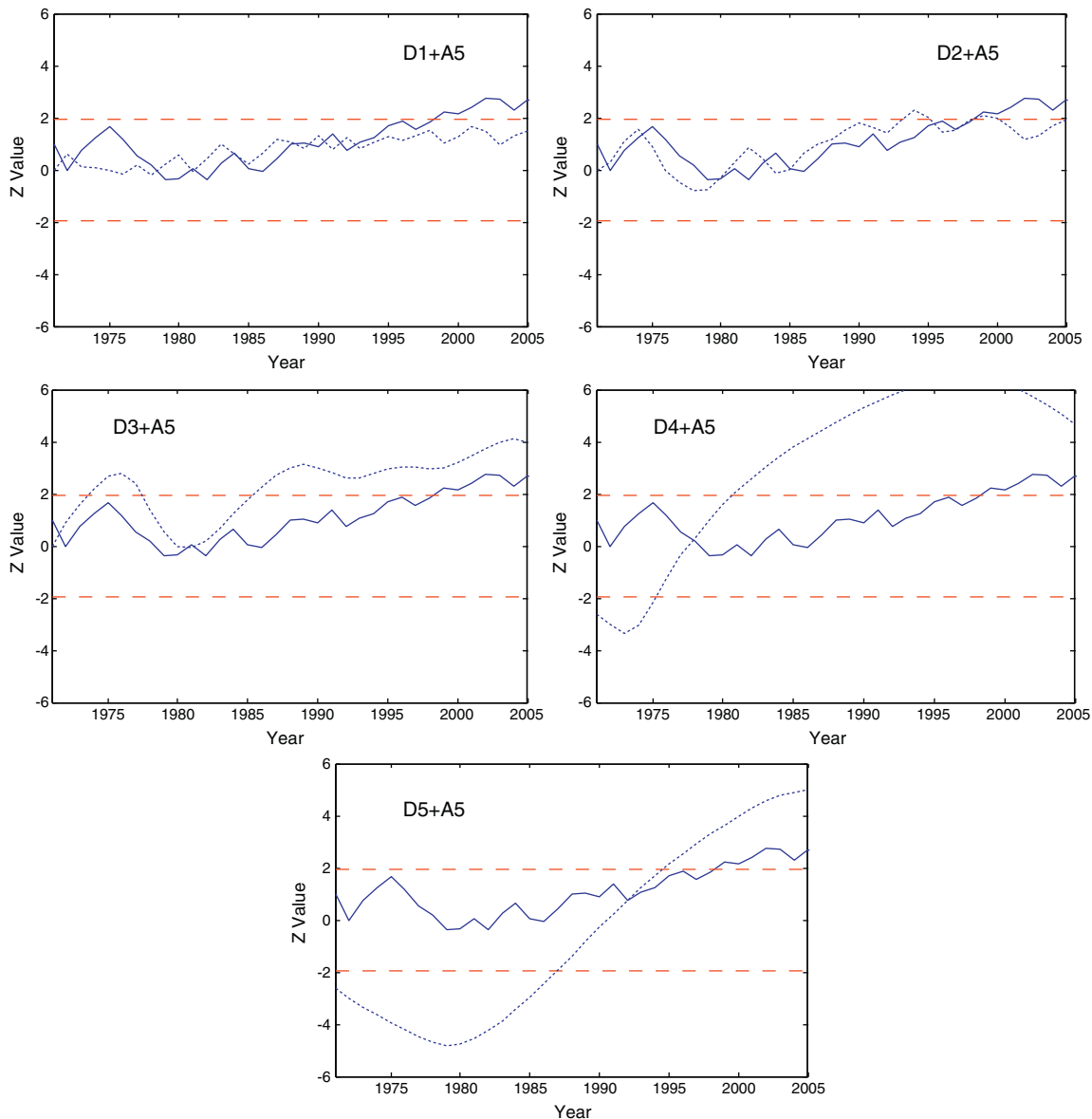
### 5.8. Autumn temperature data analysis

Although it is noted that all the MK Z-values for autumn temperatures are positive, autumn is the only season in which none of the stations explored in this study experience significant trends (Table 10). This observation is not surprising as it is in agreement with several studies where autumn has the least number of stations with significant warming (see for example: Vincent et al., 2007). It can be seen in Table 10 that the dominant periodic components affecting the trends in autumn temperature are slightly different from station to station. For stations Harrow and Peterborough, it is the D2 (4-year mode); for station Vineland, it is the D3 (8-year mode);

and for stations Belleville and Val d'Or, it is the D1 (2-year mode). Again, these important periodic modes may have coincided with some of the major peaks of the NAO cycle (i.e. 2 and 8 years). Although the NAO is not the only factor that influences the temperature variability over the study area, it can be considered an important factor.

An example of autumn temperature data decomposition and sequential MK analysis is given in Fig. 12. Generally, the trends in autumn temperatures are characterized by 2- to 8-year periodicities (Table 10), which are inconsistent with the results obtained from annual, winter, and summer data analyses where most trends are influenced by periodicities that are greater than eight years. This observation could suggest





**Fig. 11.** Sequential Mann–Kendall graphs of station Harrow's summer temperature data. The progressive trend lines of the original data are represented by the solid lines and the trend lines of the detail components (with their approximation added) are represented by the dashed lines. The upper and lower dashed lines represent the confidence limits ( $\alpha = 5\%$ ). For this station, D2 and D3 components were chosen to be the most dominant periodicities for trends.

that autumn has minimal contribution towards the warming trends observed in the annual temperature data over the study area.

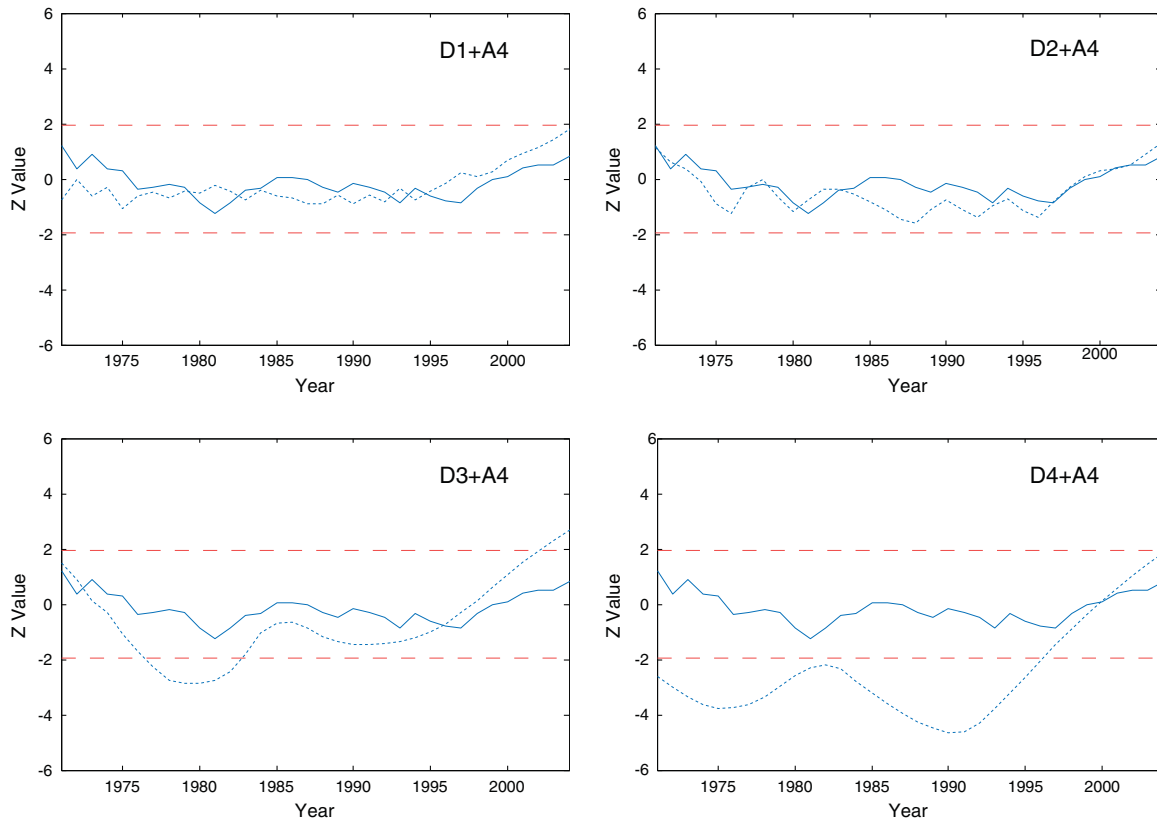
## 6. Conclusions and recommendations

Surface air temperature trends from a total of five stations located in Ontario and Quebec were analyzed using the WT and the MK trend test. The use of the DWT prior to applying the MK test in analyzing temperature trends was found to be very useful in this study. The original data were decomposed into a series of their detail and approximation components and then they were tested with the MK test. By doing so, we were able to obtain information about the periodic mode(s)

considered important in affecting the observed trends of a specific dataset.

The observations and findings of this study reveal that southern parts of Ontario and Quebec are experiencing warming trends in temperature. The use of monthly and seasonally based data in our study was found to be useful in determining the influence of higher-frequency events (short-term periodic modes) on the observed trends. Warming trends in the monthly data are affected by high-frequency periodicities ranging from 2 to 4 months, which may have masked the effect of the longer-time scales (lower frequency components). Annual periodicities are found to affect the trends in seasonally-based data.

The analysis of the lower-resolution data (i.e. annual, winter, spring, summer, and autumn) revealed that low-



**Fig. 12.** Sequential Mann–Kendall graphs of station Harrow's autumn temperature data. The progressive trend lines of the original data are represented by the solid lines and the trend lines of the detail components (with their approximation added) are represented by the dashed lines. The upper and lower dashed lines represent the confidence limits ( $\alpha = 5\%$ ). For this station, the D2 component, which represents the 4-year time mode, was considered the most dominant periodicity for trends.

frequency periodicities were more dominant in affecting the temperature trends. For annual data, the most important periodic modes that affect their trends are made up of multi-year and decadal events. Wu et al. (2007) demonstrated the superior performance of the multi-decadal trend model in capturing the variability and change in the annual global surface air temperature anomaly (GSTA) for the period 1961–1990. The rate of change in temperature using the multi-decadal trend model was higher compared to the other models, but the longer time scales were found to be more reliable when assessing trends of GSTA. Similarly, the results of the annual data analysis in the present study also revealed that the trends were affected by the higher time-scale components mostly between 8 and 16 years. These time scales explain the variability associated with the annual surface air temperature over southern Quebec and Ontario.

From the seasonal point of analysis, winter and summer are experiencing the most uniform trends in temperature, where all the sites are experiencing significant positive trends (except for Val d'Or's winter data). The results of the winter and summer analysis are also the most consistent with those of the annual data, in which most of the dominant periodic components affecting the trends are also between 8 and 16 years. Based on the findings of this study, it can be suggested that long-term trends in the temperature data over

southern Quebec and Ontario may be due to winter and summer warming. Some possible causes of winter and summer warming that have been identified in previous studies are: increases in dewpoint, air moisture, and humidity (Vincent et al., 2007), and the influence of urban heat island effects (Prokoph and Patterson, 2004).

The relationships between the temperature trends over the study area and large-scale climate circulations important for Canadian climate (e.g. the NAO, ENSO) can be quantified in future studies using correlation analysis. The findings of the present study have established the baseline information about the important periodicities that affect the temperature trends; this can then be incorporated in the future when analyzing the linkages between temperature trends in southern Ontario and Quebec, and different climatic phenomena.

Finally, it would also be very useful to include more stations (perhaps with longer data records) in future studies to obtain more representative results for the whole provinces. However, suitable interpolation methods to fill the missing records found in many stations have to be carefully determined in order to minimize the errors associated with interpolations. In this study, only five stations were included because they are the only ones in our study area that have complete records (with no missing values) for 40 years.

## Acknowledgments

This study was funded by a NSERC Discovery Grant held by Jan Adamowski, as well as a Liliane and David M. Stewart Foundation Scholarship held by Deasy Nalley.

## References

- Adamowski, K., Bougadis, J., 2003. Detection of trends in annual extreme rainfall. *Hydrol. Processes* 17 (18), 3547–3560.
- Adamowski, J., Prokoph, A., 2013. Assessing the impacts of the urban heat island effect on streamflow patterns in Ottawa, Canada. *J. Hydrol.* 496, 225–237.
- Adamowski, K., Prokoph, A., Adamowski, J., 2009. Development of a new method of wavelet aided trend detection and estimation. *Hydrol. Processes* 23 (18), 2686–2696.
- Antcl, F., Coulibaly, P., 2004. Wavelet analysis of the interannual variability in southern Québec streamflow. *J. Clim.* 17 (1), 163–173.
- Baliunas, S., Frick, P., Sokoloff, D., Soon, W., 1997. Time scales and trends in the central England temperature data (1659–1990): a wavelet analysis. *Geophys. Res. Lett.* 24 (11), 1351–1354.
- Bonsal, B., Shabbar, A., 2011. Large-scale climate oscillations influencing Canada, 1900–2008. Canadian Biodiversity: Ecosystem Status and Trends 2010. Technical Thematic Report No. 4. Canadian Councils of Resource Ministers, Ottawa, ON (iii + 15 p. <http://www.biodivcanada.ca/default.asp?lang=En&n=137E1147-0> [accessed 25 June 2012]).
- Bonsal, B.R., Prowse, T.D., Duguay, C.R., Lacroix, M.P., 2006. Impacts of large-scale teleconnections on freshwater-ice break/freez-up dates over Canada. *J. Hydrol.* 330 (1–2), 340–353.
- Bukovsky, M.S., 2012. Temperature trends in the NARCCAP regional climate models. *J. Clim.* 25, 3985–3991 <http://dx.doi.org/10.1175/JCLI-D-11-00588.1>.
- Burn, D.H., Hag Elnur, M.A., 2002. Detection of hydrologic trends and variability. *J. Hydrol.* 255 (1–4), 107–122.
- Chaoche, K., Neppel, L., Dieulin, C., Pujol, N., Ladouche, B., Martin, E., Salas, D., Caballero, Y., 2010. Analyses of precipitation, temperature and evapotranspiration in a French Mediterranean region in the context of climate change. *C. R. Geosci.* 342 (3), 234–243.
- Chou, C.-M., 2007. Applying multi-resolution analysis to differential hydrological grey models with dual series. *J. Hydrol.* 332 (1–2), 174–186.
- Cook, E.R., D'Arrigo, R.D., Briffa, K.R., 1998. A reconstruction of the North Atlantic Oscillation using tree-ring chronologies from North America and Europe. *Holocene* 8 (1), 9–17.
- Craigmile, P.F., Guttorp, P., Percival, D.B., 2004. Trend assessment in a long memory dependence model using the discrete wavelet transform. *Environmetrics* 15 (4), 313–335.
- Damyranov, N.N., Matthews, H.D., Mysak, L.A., 2012. Observed decreases in the Canadian outdoor skating season due to recent winter warming. *Environ. Res. Lett.* 7 (1), 014028.
- de Artigas, M.Z., Elias, A.G., de Campra, P.F., 2006. Discrete wavelet analysis to assess long-term trends in geomagnetic activity. *Phys. Chem. Earth.* 31 (1–3), 77–80.
- de Jager, C., Duhau, S., van Geel, B., 2010. Quantifying and specifying the solar influence on terrestrial surface temperature. *J. Atmos. Sol.-Terr. Phys.* 72, 926–937.
- Dietz, E.J., Killeen, T.J., 1981. A nonparametric multivariate test for monotone trend with pharmaceutical applications. *J. Am. Stat. Assoc.* 76 (373), 169–174.
- Domroes, M., El-Tantawi, A., 2005. Recent temporal and spatial temperature changes in Egypt. *Int. J. Climatol.* 25 (1), 51–63.
- Dong, X., Nyren, P., Patton, B., Nyren, A., Richardson, J., Maresca, T., 2008. Wavelets for agriculture and biology: a tutorial with applications and outlook. *BioScience* 58 (5), 445–453.
- Drago, A.F., Boxall, S.R., 2002. Use of the wavelet transform on hydro-meteorological data. *Phys. Chem. Earth.* 27 (32–34), 1387–1399.
- Erylkin, A.D., Sloan, T., Wolfendale, A.W., 2009. Solar activity and the mean global temperature. *Environ. Res. Lett.* 4 (2009), 014006.
- Fan, X.-H., Wang, M.-B., 2011. Change trends of air temperature and precipitation over Shanxi Province, China. *Theor. Appl. Climatol.* 103 (3–4), 519–531.
- Farge, M., 1992. Wavelet transforms and their applications to turbulence. *Annu. Rev. Fluid Mech.* 24, 395–457.
- Franzke, C., 2010. Long range dependence and climate noise characteristics of Antarctic temperature data. *J. Clim.* 23, 6074–6081.
- Gough, W.A., 2008. Theoretical considerations of day-to-day temperature variability applied to Toronto and Calgary, Canada data. *Theor. Appl. Climatol.* 94 (1–2), 97–105.
- Hamed, K.H., Rao, A.R., 1998. A modified Mann–Kendall trend test for autocorrelated data. *J. Hydrol.* 204 (1–4), 182–196.
- Hasanean, H.M., 2001. Fluctuations of surface air temperature in the Eastern Mediterranean. *Theor. Appl. Climatol.* 68 (1–2), 75–87.
- Hirsch, R.M., Slack, J.R., 1984. A nonparametric trend test for seasonal data with serial dependence. *Water Resour. Res.* 20 (6), 727–732.
- IPCC, 2007. Climate change 2007: the fourth IPCC scientific assessment. In: Parry, M.L., Canziani, O.F., Palutikof, J.P., van der Linden, P.J., Hanson, C.E. (Eds.), Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Jones, P.D., Briffa, K.R., 1992. Global surface air temperature variations during the twentieth century: part 1, spatial, temporal and seasonal details. *Holocene* 2 (2), 165–179.
- Jones, P.D., Moberg, A., 2003. Hemispheric and large-scale surface air temperature variations: an extensive revision and an update to 2001. *J. Clim.* 16, 206–223.
- Jung, H.-S., Choi, Y., Oh, J.-H., Lim, G.-H., 2002. Recent trends in temperature and precipitation over South Korea. *Int. J. Climatol.* 22 (11), 1327–1337.
- Kadioglu, M., 1997. Trends in surface air temperature data over Turkey. *Int. J. Climatol.* 17 (5), 511–520.
- Kahya, E., Kalayci, S., 2004. Trend analysis of streamflow in Turkey. *J. Hydrol.* 289 (1–4), 128–144.
- Kallache, M., Rust, H.W., Kropp, J., 2005. Trend assessment: applications for hydrology and climate research. *Nonlinear Process. Geophys.* 12 (2), 201–210.
- Karaburun, A., Demirci, A., Kara, F., 2011. Analysis of spatially distributed annual, seasonal and monthly temperatures in Istanbul from 1975 to 2006. *World Appl. Sci. J.* 12 (10), 1662–1675.
- Kendall, M.G., 1975. Rank Correlation Methods. Charles Griffin, London.
- Kravchenko, V.O., Evtushevsky, O.M., Grytsai, A.V., Milinevsky, G.P., 2011. Decadal variability of winter temperatures in the Antarctic Peninsula region. *Antarct. Sci.* 23 (6), 614–622.
- Labat, D., Goddérès, Y., Probst, J.L., Guyot, J.L., 2004. Evidence for global runoff increase related to climate warming. *Adv. Water Resour.* 27 (6), 631–642.
- Lassen, K., 1991. Long-term variations in solar activity and their apparent effect on the earth's climate. <http://www.tmgnow.com/repository/solar/lassen1.html> (accessed 25 July 2012).
- Lau, K.M., Weng, H., 1995. Climate signal detection using wavelet transform: how to make a time series sing. *Bull. Am. Meteorol. Soc.* 76 (12), 2391–2402.
- Lim, Y.-H., Lye, L.M., 2004. Wavelet analysis of tide-affected low streamflows series. *J. Data Sci.* 2, 149–163.
- Lindsay, R.W., Percival, D.B., Rothrock, D.A., 1996. The discrete wavelet transform and the scale analysis of the surface properties of sea ice. *IEEE Trans. Geosci. Remote Sens.* 34 (3), 771–787.
- Lu, Q., Lund, R., Seymour, L., 2005. An update of U.S. temperature trends. *J. Clim.* 18, 4906–4914.
- Ludwig, W., Serrat, P., Cesmat, L., Garcia-Estevés, J., 2004. Evaluating the impact of the recent temperature increase on the hydrology of the Têt River (Southern France). *J. Hydrol.* 289 (1–4), 204–221.
- Lund, R., Seymour, L., Kafadar, K., 2001. Temperature trends in the United States. *Environmetrics* 12 (7), 673–690.
- Ma, J., Xue, J., Yang, S., He, Z., 2003. A study of the construction and application of a Daubechies wavelet-based beam element. *Finite Elem. Anal. Des.* 39, 965–975.
- Makokha, G.L., Shisanya, C.A., 2010. Trends in mean annual minimum and maximum near surface temperature in Nairobi city, Kenya. *Adv. Meteorol.* 6. <http://dx.doi.org/10.1155/2010/676041> (Article ID 676041).
- Mann, H.B., 1945. Nonparametric tests against trend. *Econometrica* 13 (3), 245–259.
- Mekis, É., Vincent, L.A., 2011. An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada. *Atmosphere-Ocean* 49 (2), 163–177.
- Mimikou, M.A., Baltas, E., Varanou, E., Pantazis, K., 2000. Regional impacts of climate change on water resources quantity and quality indicators. *J. Hydrol.* 234 (1–2), 95–109.
- Mishra, A.K., Singh, V.P., 2010. Changes in extreme precipitation in Texas. *J. Geophys. Res.* 115 (D14), D14106.
- Moberg, A., Sonechkin, D.M., Holmgren, K., Datsenko, N.M., Karlen, W., 2005. Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data. *Nature* 433 (7026), 613–617.
- Mohsin, T., Gough, W., 2010. Trend analysis of long-term temperature time series in the Greater Toronto Area (GTA). *Theor. Appl. Climatol.* 101 (3), 311–327.
- Nalley, D., Adamowski, J., Khalil, B., 2012. Using discrete wavelet transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954–2008). *J. Hydrol.* 475, 204–228.
- Nicholls, R.J., Tol, R.S.J., 2006. Impacts and responses to sea-level rise: a global analysis of the SRES scenarios over the twenty-first century. *Philos. Trans. R. Soc. A* 364 (1841), 1073–1095.
- Nicholls, N., Gruza, G.V., Jouzel, J., Karl, T.R., Ogallo, L.A., Parker, D.E., 1996. Observed climate variability and change. In: Houghton, J.T., Filho, L.G.M., Callander, B.A., Harris, N., Kattenberg, A., Maskell, K. (Eds.), Climate

- Change 1995: The Science of Climate Change. 132–192. Cambridge University Press, Cambridge, UK.
- Partal, T., 2010. Wavelet transform-based analysis of periodicities and trends of Sakarya basin (Turkey) streamflow data. *River Res. Appl.* 26 (6), 695–711.
- Partal, T., Kahya, E., 2006. Trend analysis in Turkish precipitation data. *Hydrol. Processes* 20, 2011–2026.
- Partal, T., Küçük, M., 2006. Long-term trend analysis using discrete wavelet components of annual precipitations measurements in Marmara region (Turkey). *Phys. Chem. Earth* 31 (18), 1189–1200.
- Percival, D.B., 2008. Analysis of geophysical time series using discrete wavelet transforms: an overview. In: Donner, R.V., Barbosa, S.M. (Eds.), *Nonlinear Time Series Analysis in the Geosciences – Applications in Climatology, Geodynamics, and Solar-terrestrial Physics*. Springer, Berlin/Heidelberg.
- Pišoft, P., Kalvová, J., Brázdil, R., 2004. Cycles and trends in the Czech temperature series using wavelet transforms. *Int. J. Climatol.* 24 (13), 1661–1670.
- Polyakov, I.V., Bekryaev, R.V., Alekseev, G.V., Bhatt, U.S., Colony, R.L., Johnson, M.A., Maskhtas, A.P., Walsh, D., 2003. Variability and trends of air temperature and pressure in the Maritime Arctic, 1875–2000. *J. Clim.* 16 (12), 2067–2077.
- Popivanov, I., Miller, R.J., 2002. Similarity search over time-series data using wavelets. *Proceedings of the 18th IEEE International Conference on Data Engineering*, pp. 212–221.
- Prokoph, A., Patterson, R.T., 2004. Application of wavelet and regression analysis in assessing temporal and geographic climate variability: Eastern Ontario, Canada as a case study. *Atmosphere-Ocean* 43 (2), 201–212.
- Prokoph, A., Adamowski, J., Adamowski, K., 2012. Influence of the 11 year solar cycle on annual streamflow maxima in Southern Canada. *J. Hydrol.* 442–443, 55–62.
- Rebetez, M., Reinhard, M., 2007. Monthly air temperature trends in Switzerland 1901–2000 and 1975–2004. *Theor. Appl. Climatol.* 91 (1–4), 27–34.
- Shrestha, A.B., Wake, C.P., Mayewski, P.A., Dibb, J.E., 1999. Maximum temperature trends in the Himalaya and its vicinity: an analysis based on temperature records from Nepal for the period 1971–94. *J. Clim.* 12 (9), 2775–2786.
- Solheim, J.-K., Stordahl, K., Humlum, O., 2011. Solar activity and Svalbard temperatures. *Adv. Meteorol.* 8. <http://dx.doi.org/10.1155/2011/543146> (Article ID 543146).
- Statistics Canada, 2011. Temperature trends in Canada. *EnviroStats* 5 (1) (Catalogue No. 16-002-X).
- Su, H., Liu, Q., Li, J., 2011. Alleviating border effects in wavelet transforms for nonlinear time-varying signal analysis. *Adv. Electr. Comput. Eng.* 11 (3), 6.
- Torrence, C., Compo, G.P., 1998. A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* 79 (1), 61–78.
- Vincent, L.A., Gullett, D.W., 1999. Canadian historical and homogeneous temperature datasets for climate change analyses. *Int. J. Climatol.* 19 (12), 1375–1388.
- Vincent, L.A., van Wijngaarden, W.A., Hopkinson, R., 2007. Surface temperature and humidity trends in Canada for 1953–2005. *J. Clim.* 20 (20), 5100–5113.
- Vincent, L.A., Milewska, E.J., Hopkinson, R., Malone, L., 2009. Bias in minimum temperature introduced by a redefinition of the climatological day at the Canadian synoptic stations. *J. Appl. Meteorol. Climatol.* 48 (10), 2160–2168.
- Vonesch, C., Blu, T., Unser, M., 2007. Generalized Daubechies wavelet families. *IEEE Trans. Signal Process.* 55 (9), 4415–4429.
- Wang, N., Lu, C., 2009. Two-dimensional continuous wavelet analysis and its application to meteorological data. *J. Atmos. Ocean. Technol.* 27 (4), 652–666.
- Wang, J.Z., Wiederhold, G., Firschein, O., Xin Wei, S., 1998. Content-based image indexing and searching using Daubechies' wavelets. *Int. J. Digit. Libr.* 1 (4), 311–328.
- Wettstein, J.J., Mearns, L.O., 2002. The influence of the North Atlantic–Arctic oscillation on mean, variance, and extremes of temperature in the northeastern United States and Canada. *J. Clim.* 15 (24), 3586–3600.
- Wu, Z., Huang, N.E., Long, S.R., Peng, C.-K., 2007. On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proc. Natl. Acad. Sci. U. S. A.* 104 (38), 14889–14894.
- Yue, S., Pilon, P., Phinney, B., Cavadias, G., 2002. The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrol. Processes* 16 (9), 1807–1829.

Zhang, X., Vincent, L.A., Hogg, W.D., Niitsoo, A., 2000. Temperature and precipitation trends in Canada during the 20th century. *Atmos. Ocean* 38 (3), 395–429.



**Deasy Nalley** is currently pursuing her Ph.D. in the Department of Bioresource Engineering, McGill University, Canada. Her current research interests are in the area of trend estimation and analysis of hydrological data using wavelet transforms and the Mann–Kendall trend test.



**Dr. Jan Adamowski** is an Assistant Professor of Hydrology and Water Resources Management in the Department of Bioresource Engineering at McGill University in Canada. At McGill, he is also the Liliane and David M. Stewart Scholar in Water Resources, the Director of the Integrated Water Resources Management Program (which comprises a Master of Science program and an Online Certification Program), and the Associate Director of the Brace Centre for Water Resources Management. Dr. Adamowski's teaching and research activities revolve around statistical hydrology and integrated and adaptive water resources management.



**Dr. Bahaa Khalil** is a Postdoctoral Fellow in the Department of Bioresource Engineering at McGill University in Canada. Dr. Khalil's research activities revolve around assessment and redesign of environmental monitoring networks, water quality assessment, statistical hydrology, hydrologic and water quality modeling. He has 18 years of research experience with 50 publications in peer-reviewed journals, conference proceedings and technical reports.



**Dr. Bogdan Ozga-Zielinski** is an Assistant Professor in the Division of Protection and Development of Environment, in the Faculty of Environmental Engineering at the Warsaw University of Technology, where he lectures in hydrology. At the same time, he is the Head of the Centre of Hydrology at the Institute of Meteorology and Water Management-National Research Institute of Poland, where he is responsible for the coordination of water resources and applied hydrology research and international co-operation in hydrology with the World Meteorological Organization and other European entities. His main fields of interest are flood frequency analysis and mathematical modeling of hydrological processes.





# A new approach for regional scale interrill and rill erosion intensity mapping using brightness index assessments from medium resolution satellite images

Hossein Saadat<sup>a</sup>, Jan Adamowski<sup>b,\*</sup>, Vahid Tayefi<sup>c</sup>, Mohammad Namdar<sup>d</sup>, Forood Sharifi<sup>e</sup>, Sasan Ale-Ebrahim<sup>d</sup>

<sup>a</sup> McGill University, Macdonald Campus 21, 111 Lakeshore Road, Ste-Anne-de-Bellevue, Quebec, H9X 3V9, Canada

<sup>b</sup> Department of Bioresource Engineering, McGill University, Macdonald Campus 21, 111 Lakeshore Road, Ste-Anne-de-Bellevue, Quebec, H9X 3V9, Canada

<sup>c</sup> School of Geography, Tarbiyat Modares University, P.O. Box 13445, Tehran, Iran

<sup>d</sup> Forest, Range and Watershed Management Organization, Lashgark Road, Tehran, Iran

<sup>e</sup> Soil Conservation and Watershed Management Research Institute, P.O. Box 13445, Tehran, Iran

## ARTICLE INFO

### Article history:

Received 21 January 2013

Received in revised form 20 June 2013

Accepted 13 August 2013

### Keywords:

Interrill and rill erosion intensity

Brightness index

Landsat ETM+ imagery

Geographic information system (GIS)

Golestan dam watershed

Ancillary layers

## ABSTRACT

Having accurate soil erosion intensity/type maps using satellite imagery is not generally a difficult task. However, there are still difficulties for the generation of small scale erosion features at regional and national levels. It is even more problematic when high-resolution satellite images cannot be used due to their high cost at a regional level. The principal objective of this study is to investigate the applicability of brightness value to generate accurate interrill and rill erosion intensity maps using medium resolution satellite images at a regional level. In this study, Landsat ETM+ images are used and the Golestan dam watershed with an area of 4511.8 km<sup>2</sup> located at northeast of Iran is selected as the study area. In order to generate a Homogeneous Land Unit (HLU) map, three ancillary layers including slope, landform, land use and land cover, are overlaid on each other. The HLUs are used in a supportive role for identifying appropriate sampling points across the entire study area, at which the degrees of interrill and rill erosions are measured. The ground-truth erosion information collected at the 1328 locations is divided into training and reference data sets. Using the Tasseled Cap transformation technique, the brightness value of each pixel at the beginning (May), middle (July) and end (September) of growing season is obtained. By subtracting the May brightness value ( $B_M$ ) from the July one ( $B_J$ ), and the July brightness value from the September one ( $B_S$ ), two new brightness images representing the brightness variations over May–July ( $B_{MJ}$ ) and July–September ( $B_{JS}$ ) are created. The two new brightness images are combined to generate a map where its pixels indicate the state (i.e. increase, I, decrease, D, and constant, C) of brightness variation over the two growing seasons. Using the measured interrill and rill erosion information at the training sampling locations, a unique relationship is found between the trend of brightness variation and the erosion intensity. This relationship is validated using the reference data sets. The results show that the proposed method is able to produce an interrill–rill erosion intensity map with an overall field-checked accuracy of 96% at this study location. The main advantages of this method are its high accuracy, its lower demands on time and funds for field work, and the ready availability of required data.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

Long recognized as a serious worldwide land degradation problem, soil erosion has a strong negative impact on the environment by reducing soil productivity and increasing sediment and other pollution loads into receiving water bodies (Morgan, 2005). The importance of soil erosion type/intensity maps in natural resources, agricultural, soil conservation, land management and water resources management planning and development has been recognized for decades. Over the

past decades, many models have been introduced and used for creating soil erosion maps (Flanagan et al., 2001; Parsons and Wainwright, 2006; Renard et al., 1997). New technologies (e.g. satellite data) and increased computing power have led to the development of new models in the context of soil erosion mapping (Morgan, 2005). Remote sensing provides detailed information over large regions with a regular revisit capability, and can greatly contribute to regional erosion assessment (Siakeu and Oguchi, 2000; Vrieling et al., 2008). Satellite imagery can assist soil erosion assessment/mapping through (i) automatic identification of large scale erosion (Vrieling et al., 2007) and its consequences (Jain et al., 2002), (ii) the assessment of erosion controlling factors (King et al., 2005; Vrieling et al., 2008), and (iii) the interpretation and

\* Corresponding author. Tel./fax: +1 514 398 7786.

E-mail address: [jan.adamowski@mcgill.ca](mailto:jan.adamowski@mcgill.ca) (J. Adamowski).

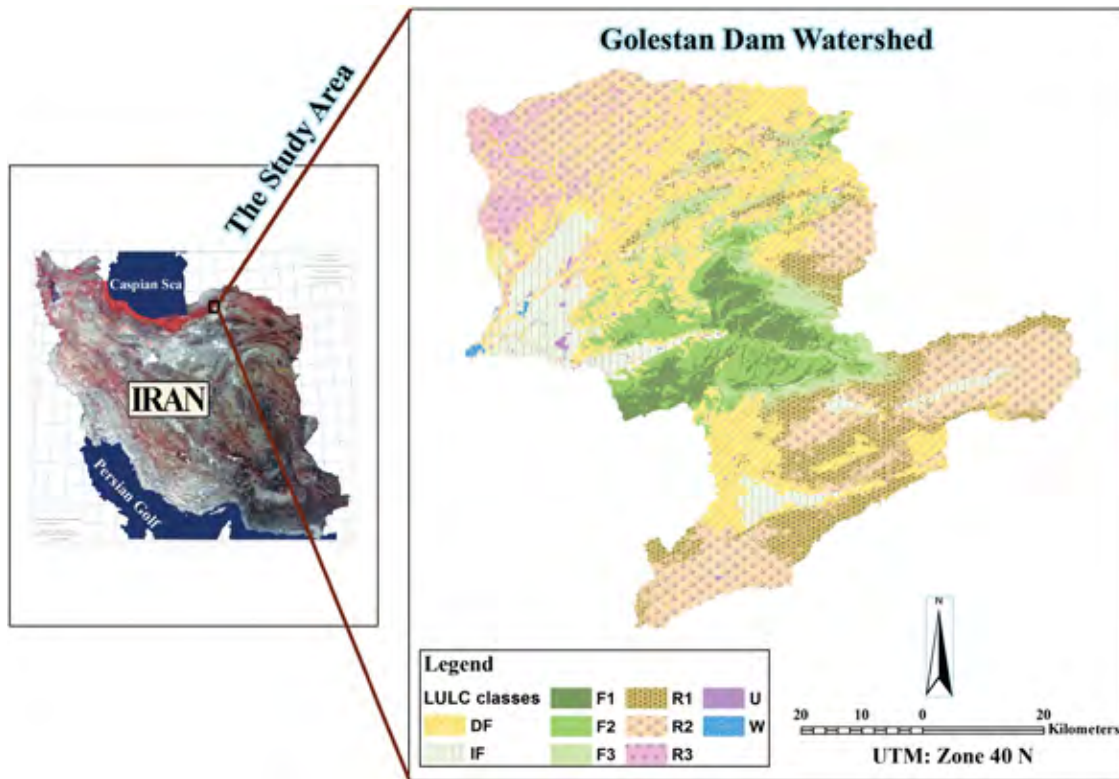


Fig. 1. Location of the study area.

classification of soil erosion using satellite data combined with additional data sources (ancillary layers).

In the first case, where high-resolution satellite data are used, large-scale erosion types (e.g. gully and badland) can be identified automatically from the satellite data. In the second case, a variety of erosion controlling factors such as vegetation, topography, climate, and soil characteristics can be obtained from satellite imagery (King et al., 2005). These data can then be used in a range of empirical models, leading to the quantitative assessment of small-scale erosion features. For instance, the empirical Universal Soil Loss Equation (USLE) can be used, in which long-term mean interrill and rill erosions are assessed (Wischmeier and Smith, 1978). Similarly, the USLE, its revised (RUSLE) and modified (MUSLE) versions (Renard et al., 1997; Smith et al., 1984), the Water Erosion Prediction Program (WEPP) (Flanagan et al., 2001), and the European Soil Erosion Model (EUROSEM) (Morgan, 1995; Morgan et al., 1984) can be used in which erosion controlling factors extracted from satellite data at basin or even regional scales play a role in assessing/mapping erosion intensities (de Vente and Poesen, 2005). The development of such models based on these factors is problematic as a number of these factors are difficult to be assessed, and are not constant in space and time and interact with each other (de Vente and Poesen, 2005). In addition, fixed data requirements, and the fact that these models are developed for a certain region, scale, or specific process and that they only provide a mean quantitative assessment of erosion phenomena, are drawbacks to their general use (Rudra et al., 1998; Vrieling, 2006). Although outputs of such models are helpful in prioritizing conservation projects within a watershed, they cannot provide detailed information of map erosion features (Poesen et al., 2003).

In the third case, erosion types/intensities can be differentiated using satellite data combined with some ancillary layers (Dymond et al., 2002; Focardi et al., 2008; Healey et al., 2005; Vrieling et al., 2008). Remote sensing data provides detailed information over large regions with a regular revisit capability, and can greatly contribute to regional erosion assessment (Siakeu and Oguchi, 2000; Vrieling et al., 2008). For

instance, Liberti et al. (2008) mapped badland areas using landsat TM/ETM satellite imagery with the aid of some morphological maps. Vrieling et al. (2008) created soil erosion risk maps using a time series of MODIS (Moderate Resolution Imaging Spectroradiometer) and ASTER (Spaceborne Thermal Emission and Reflection Radiometer) images plus a digital elevation model (DEM) and temporal rainfall data. They reported a strong relationship between the normalized difference vegetation index (NDVI) and erosion risk. However, most approaches using remotely sensed data have concentrated on mapping large-scale erosion features (e.g. gullies) and erosion risk (King et al., 2005), but little has been done with regard to creating maps showing the erosion intensity classes over the land surface by interrill and rill processes (de Vente et al., 2008). It should also be noted that medium resolution satellite data alone cannot provide appropriate information for mapping small scale erosion features like interrill and rill (Alewel et al., 2008; Vrieling, 2006). A review of existing approaches for erosion types/intensities mapping shows that although the application of high-resolution satellite images (e.g. QuickBird, GeoEye and IKONOS) may lead to very accurate soil erosion type/intensity maps, these data are not accessible for all countries (such as the study area of this project: Iran), and are very expensive if used at the regional and national scales. Thus, it can be seen that there is a gap in the literature and in practice for methods with the capability of creating soil erosion maps (particularly small scale soil erosion features) that use medium resolution satellite images at the regional and national scales (Vrieling, 2006).

The principal objective of the present study is to investigate the applicability of the brightness index (Tasseled Cap) variation over the growing season to generate accurate interrill and rill erosion intensity maps using medium resolution satellite images and some ancillary layers at a regional level.

## 2. The study area

With an area of 4511.8 km<sup>2</sup>, Iran's Golestan dam watershed is located between 55° 21' and 56° 28' E longitude, and 36° 44' and 37° 49' N

latitude, in the northeast portion of the Golestan province (Fig. 1). This sub-watershed of the Gorgan River watershed is a complex combination of mountains, hills, plains and rivers. The highest elevation is 2492 m above mean sea level and the lowest is 47 m. Because of its geographic situation and topography, a wide range of climates prevail across the different portions of the watershed; from semi-arid in the north-west and south to humid in the central portion. Mean annual precipitation ranges from 135 mm to 700 mm and mean annual air temperature from 8.5 to 17 °C. March is the month of greatest rainfall and June to October are the dry months. Different sedimentary rocks such as limestone, sandstone, shale, dolomite, and marl, along with conglomerate, loess sediments and alluvium cover the area (Lar Consulting Engineering, 2007).

Existing land use and land cover (LULC) maps show that rangeland covers an area of 200,315 ha (44.4% of the whole study area), forest an area of 98,521 ha (21.83%), dryland farming an area of 110,135 ha (24.41%), irrigated farming an area of 39,068 ha (8.66%), with urban areas contributing less than 1% (Saadat et al., 2011). Located in this area, the 92,000 ha Golestan Forest National Park is recognized by UNESCO as part of the international network of Biosphere Reserves (Japan International Cooperation Agency, 2005). Accelerated soil erosion, high sediment yields, floods and debris flows are serious problems in the Golestan dam watershed (Japan International Cooperation Agency, 2005). For example, close to 500 people were killed, thousands of livestock were lost and many infrastructures such as bridges and roads were washed out or damaged due to flooding and debris flows during the summers of 2001, 2002 and 2005 (Japan International Cooperation Agency, 2005; Sharifi et al., 2002).

### 3. Materials and methods

#### 3.1. Materials

To implement the research methodology, the following data, image and map materials are required:

- (i) Growing season Landsat ETM+ images: spring (10 May, 2003), summer (10 July, 2002), and late-summer (12 September, 2002); with 28.5 m spatial resolution in multispectral bands.
- (ii) 1:25000 digital topographic maps prepared (based on 1993 aerial photos) by the National Cartographic Center of Iran and the Forest, Range and Watershed Management Organization.
- (iii) Three digital ancillary layers to assist the production of homogeneous land units. These layers consist of a 1:25000 scale slope (LS) map (Saadat et al., 2008), a 1:50000 scale landform (LF) map (Saadat et al., 2008), and a 1:50000 scale land use and land cover (LULC) map (Saadat et al., 2011). The legend classification for the slope, landform, and land use land cover maps can be found in the footnotes on the next page.
- (iv) ERDAS Imagine (version 8.7) and ArcInfo (version 9) software are used for image processing and data analyses.

#### 3.2. Methodology

As previously mentioned, the principal objective of this research work is to map intensity of interrill and rill erosions using the brightness variation over the growing season in a large-scale watershed. To do that, five main steps are proposed (Fig. 2a, b, and c).

##### 3.2.1. Preprocessing of satellite images

The 6 bands of each ETM+ images (except thermal bands) were stacked. These images have their own map projection according to orbital parameters but to be able to compare separate images pixel by pixel, the pixel grids of each image must conform to the other images in the data base. The September stacked image using 160 ground control points taken from 1:25000 topographic maps is geometrically corrected. The total RMS Error (distance in pixel widths) in this stage

was equal to 0.8 and the first polynomial order was used in order to have less distortions. At the next stage the other two stacked images are registered to the September image (Fig. 2a. operations 1 and 2).

##### 3.2.2. Ground-truth sampling locations

This step aims to determine an appropriate number and location of sampling points over the study area, at which the degree of interrill and rill erosions is measured. The number and location of the sampling points should be specified in a way that entails the full range of erosion intensities in any possibility of land configuration (e.g. slope, land use and landform). To do that, three ancillary layers, including slope with 9 classes,<sup>1</sup> landform with 7 classes<sup>2</sup> and land use land cover with 8 classes,<sup>3</sup> is overlaid on each other in order to generate a Homogeneous Land Unit (HLU) map (Fig. 2c, operations 8 and 9). In this case study, the HLU map contains 147 unique units. Since each driven unit has similar characteristics of erosion controlling factors (similar land form, slope, land use and vegetation density), this map was used in an important supportive role for identifying appropriate sampling locations across the entire study area (de Vente and Poesen, 2005).

Using the stratified random sampling technique (Stehman, 1999), a total of 1328 ground control points over the entire study area were initially extracted. Nineteen of these sites are inaccessible due to physical barriers or remoteness from roads, and inevitably omitted from the sampling points list. These sites are replaced by 19 additional accessible sites with the same attributes. Erosion type information collected at the 1328 locations are divided into training and reference data sets. The training data set consists of the 50% collected points (664 points) that are used in the training process and the rest are treated for accuracy assessment process of soil erosion intensity classification.

##### 3.2.3. Field assessment of interrill and rill erosions

This step aims to collect the ground-truth interrill and rill erosion severity data at the sampling locations. A topographic map (1:25000) and a GPS unit are used to locate each site. Based upon the handbook provided by Stocking and Murnaghan (2001) for the field assessment of land degradation, a number of indicators are selected to determine the degree of interrill and rill erosions in an area around the sampling location equal to 1 to 3 image pixels. Since ETM+ multispectral bands were used in this study, then based on their pixel size, the plot area was considered to be approximately between 30 \* 30 and 90 \* 90 m (Fig. 2c, operation 10). These severity indicators include: a) soil color condition, b) plant and tree root exposure, c) build-up against tree trunk/plant stem of soil and/or plant debris, d) rock outcrops or stony soil surface, e) presence, depth and length of rills, and f) distance between rills.

In order to establish a measure of interrill and rill erosion severity, Table 1 is generated using the considered severity indicators and a combination of the methods provided by FAO (2006), Morgan (2005), and Stocking and Murnaghan (2001). According to the collected ground-truth data at the sampling locations, fifteen erosion types are identified, as illustrated in the matrix of Table 2 (Fig. 2c, operation 11).

##### 3.2.4. Assessment of brightness index variation

As previously mentioned, the main objective of this research is to determine interrill and rill erosion intensity using the brightness index variation over a growing season. Therefore, the Tasseled Cap transformation technique is initially performed on each image to determine

<sup>1</sup> Slope classes include <2%, 2–5%, 5–8%, 8–12%, 12–15%, 15–25%, 25–40%, 40–60%, and ≥60%.

<sup>2</sup> Landform classes contain River Alluvial Plains (RP), Piedmont Plains (PD), Gravelly Fans (GF), Upper Terraces (TRu), River Terraces (TRr), Hills (H) and Mountains (M).

<sup>3</sup> Land use and land cover classes include Irrigated Farming (IF), Dryland Farming (DF), high density Forest (F1, cover > 70%), medium density Forest (F2, 40% < cover ≤ 70%), low density Forest (F3, cover ≤ 40%), high density Rangeland (R1, cover > 30%, mostly between 30% and 50%), medium density Rangeland (R2, 15 < cover ≤ 30%), and low density Rangeland (R3, cover ≤ 15%).



the value of brightness index (Fig. 2b. operations 4, 5, 6, and 7). The technique reduces the six ETM+ reflectance bands of a single image to three individual indices termed brightness, greenness, and wetness. These three individual indices are derived using at-satellite reflectance based coefficients (Huang et al., 2002). The use of the DN (digital number) based transformation in multiscene applications can be problematic, because changing sun illumination geometry strongly affects DN, and thus affects the derived Tasseled Cap value. However, a large part of the impact of illumination geometry can be normalized by converting DN to at satellite reflectance. Therefore, a transformation based on at-satellite reflectance is more appropriate for regional applications where atmospheric correction is not feasible (Crist and Kauth, 1986).

The difference in the values of brightness is computed for each pixel by subtracting the values of one imaging date from the next. To be

specific, the May brightness value is subtracted from the July one, termed as the  $B_{MJ}$  image, and the July brightness value is subtracted from the September one, termed as the  $B_{JS}$  image. In each interval period, the difference in the value of brightness of each pixel may decrease, increase, or remain unchanged, termed as D, I and C, respectively. Hence, the attribute of each pixel in the  $B_{MJ}$  and  $B_{JS}$  images can take terms D, I and C.

The  $B_{MJ}$  and  $B_{JS}$  images are combined to get a unique attribute table representing the trend of the brightness index variation over the May–July and July–September intervals, respectively. The resultant image contains 7 classes including DD (the brightness values decreases over both interval periods), DI (the brightness values decreases over the May–July interval and increases over the July–September interval), DC (the brightness values decreases over the first interval period and that

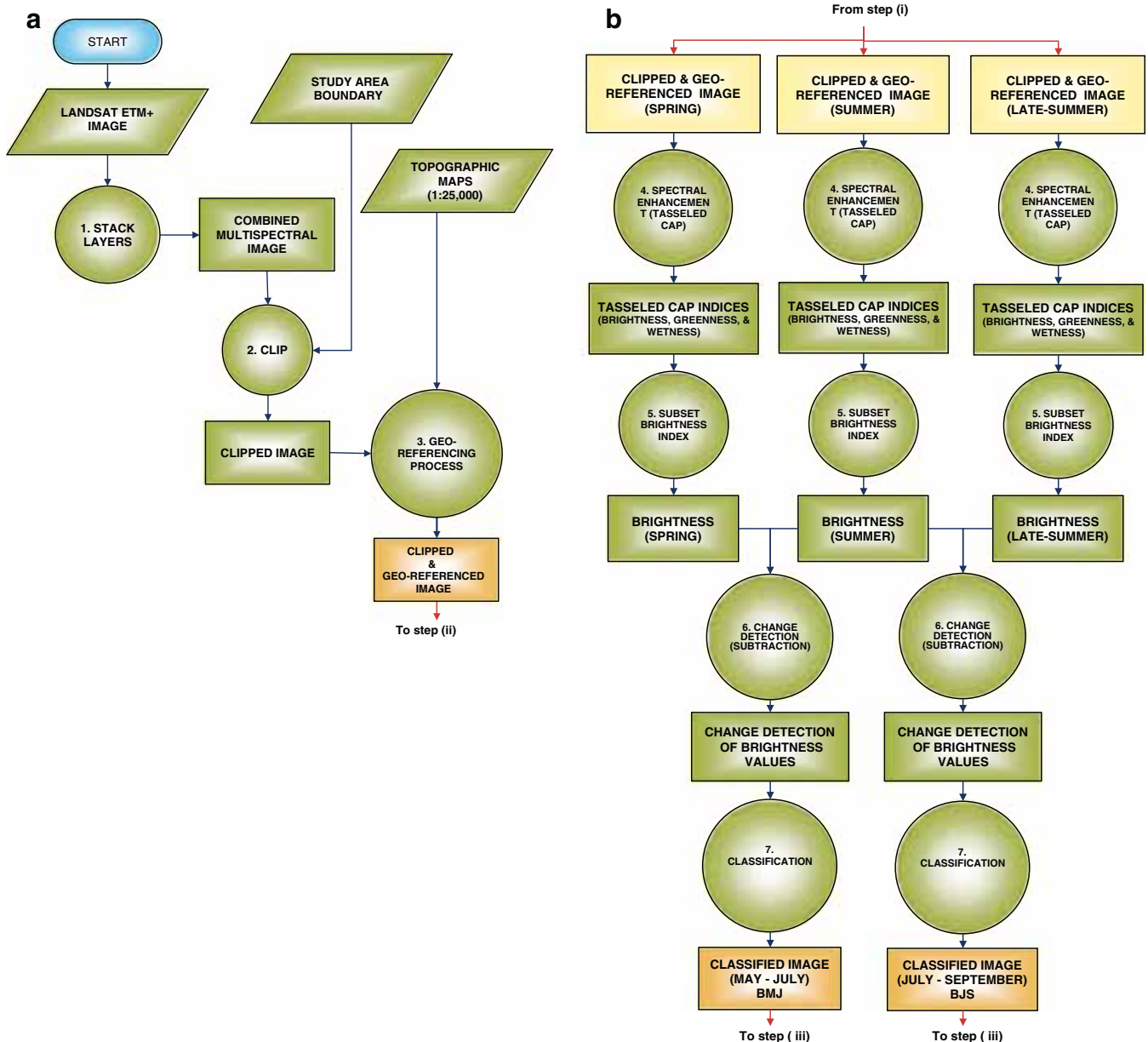


Fig. 2. Processing steps for (a) preprocessing of the images, (b) applying Tasseled Cap transformation, (c) extracting homogeneous land units and sampling location map, field data collection, analysis and classification, extending and identifying soil erosion intensities. Circles = operation, trapezoids = data input, and rectangles = maps or coverages.



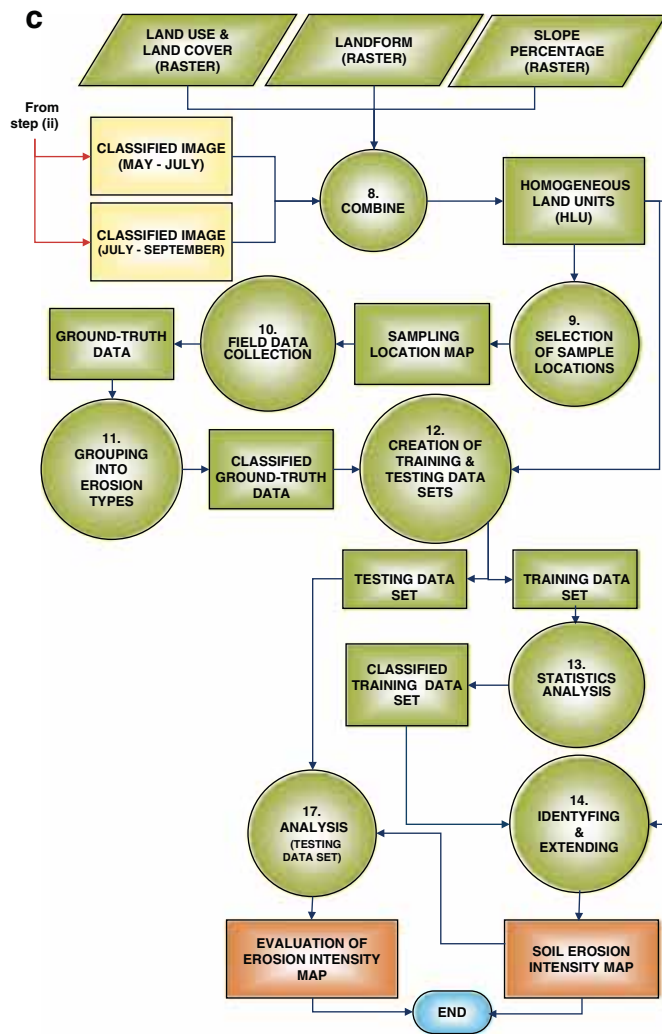


Fig. 2 (continued).

the brightness values remains unchanged over the second interval), II, IC, CI and CC.

### 3.2.5. Brightness changes and erosion intensity relationship

In order to increase the chance to make a logical relation between the trend of brightness index variation and interrill–rill erosion types, the 15 interrill–rill erosion types are subsequently grouped into five categories (five erosion intensities<sup>4</sup>) as shown in Table 2. The five categories comprise slight erosion intensity ( $E_S$ ), moderate erosion intensity ( $E_M$ ), high erosion intensity ( $E_H$ ), severe erosion intensity ( $E_{SE}$ ), and very severe erosion intensity ( $E_{VSE}$ ). Based on the collected ground-truth data at the sampling locations, a database is created for all 1328 field sites, such that each site is assigned one of five interrill–rill erosion intensities (Table 3).

Using the erosion intensities and the  $B_{MJ}$  and  $B_{JS}$  trends at the training sampling points, a logical relationship is found between them

<sup>4</sup> Slight erosion intensity ( $E_S$ , less than 5% of soil area affected by interrill and rill erosion), moderate erosion intensity ( $E_M$ , 5 to 10% of soil area affected by interrill and rill erosion), high erosion intensity ( $E_H$ , 10 to 25% of soil area affected by interrill and rill erosion), severe erosion intensity ( $E_{SE}$ , 25 to 50% of soil area affected by interrill and rill erosion) and very severe erosion intensity ( $E_{VSE}$ , more than 50% of soil area affected by interrill and rill erosion).

(Table 3). Based on this relationship, the erosion intensity of each HLU polygon can be derived based on the trend of the brightness value variation and zonal analysis. For instance, in Table 3, one can see that given the trend of the brightness index variation over the two growing season intervals (i.e.  $B_{MJ}$  and  $B_{JS}$ ) is DI, one can expect high erosion intensity (i.e.  $E_H$ ) at the HLU of interest. To assess the accuracy of the established relationship, it is applied on all HLUs over the study area to derive the erosion intensity at each HLU.

## 4. Accuracy assessments

Using the error matrix method (Congalton, 1991), the accuracy of the method is assessed using a different set of 50% referenced points. This is done only in forest and rangeland areas (Table 4). As mentioned earlier, the capability of the proposed method is only assessed in forest (F) and rangeland (R) land uses since the established logical relationship between the erosion intensities and the  $B_{MJ}$  and  $B_{JS}$  trends is not found in other land uses due to anthropogenic effects in them.

## 5. Results and discussion

Considering only rangeland (R) and forest (F) land uses, a logical relationship exists between the interrill–rill erosion intensity of each HLU and the trend of the brightness index variation. For example, as seen in Table 3, in the case of the slight interrill–rill erosion intensities ( $E_S$ ) the trend of the brightness index variation over the two growing season intervals is DD, excluding rocky areas of more than 80% outcrop ( $RO_D$ ) which is CC. In the case of  $E_M$ ,  $E_H$ ,  $E_{SE}$  and  $E_{VSE}$  the trends of the brightness index variation are DC, DI, CI and II, respectively. This means that by having the brightness index trend change over the two growing season intervals, interrill–rill erosion intensities can be determined at each HLU where land use is either forest or rangeland. Having determined interrill–rill erosion intensities, an interrill–rill soil erosion intensity map can then be generated for the study area (Fig. 3).

Such a direct relationship does not exist for other land uses (i.e. IF and DF), most likely due to a masking effect associated with the greater degree of anthropogenic activities within these land uses. Irrigated farming (IF) and dryland farming (DF) both experience much more anthropogenic land alteration than rangeland (R) and forest (F). This anthropogenic alteration of IF and DF tends to mask, or remove to a large degree, evidence of rill and interrill erosions, and so changes in brightness are not a result of natural rill or interrill development but of anthropogenic alterations.

Considering only R and F and excluding  $RO_D$  land uses, for the lower erosion intensities ( $E_S$ ,  $E_M$  and  $E_H$ ), all have a brightness trend of D in the first interval. As erosion intensities move from  $E_S$  to  $E_M$  and then to  $E_H$ , the brightness trend goes from D to C, and then to I in the second interval, respectively. Even though spring rains cause the development of rill and interrill features, increasing vegetation growth in the early season masks this and the net impact is a brightness trend of D in the first interval for all three lower erosion intensities ( $E_S$ ,  $E_M$  and  $E_H$ ). Subsequently, as summer progresses with little rain, vegetative cover decreases. For  $E_S$  where less than 5% of the HLU area experiences erosion, vegetation is still dense enough to mask the erosion and the brightness trend continues as D. For  $E_M$  where between 5 and 10% of the HLU area is eroded, vegetative cover is less effective and the brightness trend becomes C. Finally, for  $E_H$  where 10 to 25% the HLU area is eroded, vegetative cover is even less effective and the brightness trend changes to I. As such, for the lower erosion intensity ranges, it is the brightness trend in the second interval of the growing season ( $B_{JS}$ ) which determines the degree of erosion intensity among  $E_S$ ,  $E_M$ , and  $E_H$ .

For the areas of severe and very severe erosion intensity (i.e.  $E_{SE}$  and  $E_{VSE}$ ), changes in the brightness trend is also evident. In the case of  $E_{SE}$  and  $E_{VSE}$ , the brightness trend in the first interval is C and I, respectively, while the brightness trend in the second interval for both cases remains unchanged as I. In these cases, vegetative cover masks erosion

**Table 1**  
Means of establishing severity of interrill and rill erosion.

Feature of erosion	Degree	Description
Interrill erosion (S)	Not apparent (S0) Slight (S1)	No obvious signs of interrill erosion. No evidence of plant root exposure but in less than 5% of the area plant debris or/and topsoil particles removed from their original location through surface wash. Soil level slightly higher on upslope of plants and boulders. At some locations the surface is covered by more than 80% rock outcrops (RO <sub>D</sub> ) and only less than 5% of the soil surface is affected by interrill erosion (S1 + RO <sub>D</sub> ).
	Moderate (S2)	Plant debris or/and topsoil particles removed from their original location through surface wash in 5 to 10% of the area. Most of these particles are deposited on the upslope side of plant stems with a height of 5–10 mm and plant roots are partially exposed above the present soil surface with a height of 5–20 mm. S2 is usually restricted to areas with no rill to high rill erosion, depending on the land slope. At some locations the surface is covered by 40–80% rock outcrops (RO <sub>A</sub> ) and just 5 to 10% of the soil surface is affected by interrill erosion (S2 + RO <sub>A</sub> ).
	High (S3)	Plant debris and topsoil particles removed through surface wash in 10 to 25% of the area. Heavier particles are deposited on the upslope side of plant stems with a height of 10–30 mm and plant roots are partially exposed above the present soil surface with a height of 20–50 mm. S3 is usually restricted to areas with moderate to severe rill erosion (Ri2, Ri3, or Ri4).
	Severe (S4)	Topsoil particles transported through surface wash in 25 to 50% of the area. Heavier particles are deposited on the upslope side of plant stems with a height of 30–50 mm and plant roots are partially exposed above the present soil surface over 50 mm high. S4 is usually restricted to areas with high to very severe rill erosion (Ri3 or Ri5).
	Very severe (S5)	Most (>50%) of the original soil surface removed and subsoil horizons exposed at or close to the soil surface. Accumulation of soil on the upslope side of plants and boulders are over 50 mm high. Extensive exposure of plant roots occurred in more than 50% of the area. S5 is usually restricted to areas with severe rill erosion (Ri4).
Rill erosion (Ri)	Not apparent (Ri0) Slight (Ri1)	No rills present. A few shallow rills, less than 10 cm depth spaced every 20–50 m, which are able to transport soil particles from their original location. Ri1 is always associated with S1 or S2.
	Moderate (Ri2)	Discontinuous rills, less than 15 cm depth spaced every 10–20 m. Plant debris and soil particles are transported via runoff. Ri2 is usually associated with S2, and in some areas where slope is over 8%, with S3.
	High (Ri3)	A continuous network of shallow to moderately deep rills (less than 20 cm depth) spaced every 5–10 m. Soil particles, even heavier particles are transported and deposited long distances. Ri3 is usually associated with S2, S3, or S4 depending on vegetation cover and slope.
	Severe (Ri4)	A continuous network of deep rills (up to 30 cm depth) spaced every 2–5 m. Most of the original soil surface is removed. Ri4 is always associated with S3, S4 or S5 depending on LULC and slope.
	Very severe (Ri5)	An extensive network of rills spaced less than 2 m. Most of the original soil surface removed. Ri5 is usually associated with S4.

development to a lesser degree compared to lower erosion intensities. As seen in Table 3, the brightness trend in the first interval of the growing season (B<sub>MJ</sub>) is a determinant of E<sub>SE</sub> and E<sub>VSE</sub>.

**Table 2**  
Grouping interrill and rill erosions into erosion type and intensity.

Rill \ Interrill	Ri0 (Not apparent)	Ri1 (Slight)	Ri2 (Moderate)	Ri3 (High)	Ri4 (Severe)	Ri5 (Very severe)
S0 (Not apparent)	-	-	-	-	-	-
S1 (Slight)	S1 (E <sub>S</sub> )	S1Ri1 (E <sub>S</sub> )	-	-	-	-
S1+RO <sub>D</sub> (Slight)	-	S1Ri1+RO <sub>D</sub> (E <sub>S</sub> )	-	-	-	-
S2 (Moderate)	S2 (E <sub>M</sub> )	S2Ri1 (E <sub>M</sub> )	S2Ri2 (E <sub>M</sub> )	S2Ri3 (E <sub>H</sub> )	-	-
S2+RO <sub>A</sub> (Moderate)	-	S2Ri1+RO <sub>A</sub> (E <sub>M</sub> )	S2Ri2+RO <sub>A</sub> (E <sub>M</sub> )	-	-	-
S3 (High)	-	-	S3Ri2 (E <sub>H</sub> )	S3Ri3 (E <sub>H</sub> )	S3Ri4 (E <sub>SE</sub> )	-
S4 (Severe)	-	-	-	S4Ri3 (E <sub>SE</sub> )	-	S4Ri5 (E <sub>VSE</sub> )
S5 (Very severe)	-	-	-	-	S5Ri4 (E <sub>VSE</sub> )	-

E <sub>S</sub> :	Slight erosion intensity (less than 5% of soil area affected by erosion)	E <sub>M</sub> :	Moderate erosion intensity (5 to 10% of soil area affected by erosion)	E <sub>H</sub> :	High erosion intensity (10 to 25% of soil area affected by erosion)
E <sub>SE</sub> :	Severe erosion intensity (25 to 50% of soil area affected by erosion)	E <sub>VSE</sub> :	Very severe erosion intensity (more than 50% of soil area affected by erosion)		

RO<sub>A</sub>: surface cover of rock outcrops is 40–80%.  
RO<sub>D</sub>: surface cover of rock outcrops is more than 80%

Using 531 referenced sites and the method of Congalton (1991), the interrill–rill erosion intensity for each of the testing ground-truthing points was evaluated (Table 4). An overall accuracy of 96% was achieved. The producer’s accuracy (PA) is a measure of how correct the classification is. As shown in Table 4, the PA ranged from a low of 92.3% in the case of E<sub>SE</sub> to a high of 98% in the case of E<sub>S</sub>. The user’s accuracy (UA) is a measure of how well the classification process captures all occurrences of any of the five erosion intensity types. In this study, the UA ranged from a low of 92.3% in the case of E<sub>SE</sub> to a high of 98% in the case of E<sub>M</sub>.

Based on the proposed method, the interrill–rill erosion intensities of all HLUs across the study area using the trend of brightness index

**Table 3**  
The 15 erosion types and their corresponding descriptive combinations of B<sub>MJ</sub> and B<sub>JS</sub>.

Erosion types <sup>1</sup>	B <sub>MJ</sub> <sup>2</sup>	B <sub>JS</sub> <sup>2</sup>	E <sup>3</sup>
S1	D	D	E <sub>S</sub>
S1Ri1	D	D	
S1Ri1 + RO <sub>D</sub> <sup>4</sup>	C	C	
S2	D	C	E <sub>M</sub>
S2Ri1	D	C	
S2Ri2	D	C	
S2Ri1 + RO <sub>A</sub> <sup>4</sup>	D	C	
S2Ri2 + RO <sub>A</sub>	D	C	
S2Ri3	D	I	E <sub>H</sub>
S3Ri2	D	I	
S3Ri3	D	I	
S3Ri4	C	I	E <sub>SE</sub>
S4Ri3	C	I	
S4Ri5	I	I	E <sub>VSE</sub>
S5Ri4	I	I	

<sup>1</sup> The 15 erosion types as determined by ground-truth data.

<sup>2</sup> B<sub>MJ</sub>: classes derived by subtracting May (spring) brightness from July (summer) brightness and B<sub>JS</sub>: classes derived by subtracting July brightness from September (late-summer) brightness; with the result that some of these values decreased (D), some increased (I), and the others remained without changing (C).

<sup>3</sup> E: interrill and rill erosion intensities, E<sub>S</sub>: slight erosion intensity, E<sub>M</sub>: moderate erosion intensity, E<sub>H</sub>: high erosion intensity, E<sub>SE</sub>: severe erosion intensity, and E<sub>VSE</sub>: very severe erosion intensity.

<sup>4</sup> RO<sub>D</sub>: surface cover of rock outcrops is more than 80% and RO<sub>A</sub>: surface cover of rock outcrops is 40–80%.

**Table 4**  
Testing data set error matrix of classification process.

Interrill–rill erosion intensities	Ground-truth data					New classified	User's accuracy %
	E <sub>S</sub>	E <sub>M</sub>	E <sub>H</sub>	E <sub>SE</sub>	E <sub>VSE</sub>		
E <sub>S</sub>	150	7	0	0	0	157	95.5
E <sub>M</sub>	3	195	1	0	0	199	98
E <sub>H</sub>	0	0	68	3	0	71	95.8
E <sub>SE</sub>	0	0	3	60	2	65	92.3
E <sub>VSE</sub>	0	0	0	2	37	39	94.9
Total sites visited	153	202	72	65	39	531	
Producer's accuracy %	98	96.5	94.4	92.3	94.9		
Overall classification accuracy = 96%							

change over the two intervals of growing season is determined. This means that by applying this method over the study area, an interrill–rill erosion intensity map is created. To further verify the credibility of this method in creating such a map, the consistency between the attributes (i.e. land use, land cover and slope) belonging to each class of erosion intensity over the whole study site is studied. To do that, these attributes are firstly extracted from the produced soil erosion intensity map (see Table 5). As seen in Table 5, as erosion intensity increases (from E<sub>S</sub> toward E<sub>VSE</sub>), land use changes from high density forests to poor rangelands. Similarly, as erosion intensity increases (from E<sub>S</sub> toward E<sub>VSE</sub>), land slope increases except in high and medium density forests (F1 and F2). F1 and F2 are correctly classed as erosion intensity type E<sub>S</sub>, even in areas of steep and very steep slopes. These shows that there is a consistency between the attributes of each class of erosion intensity, and corroborates other research (de Vente et al., 2008; Kimaro et al., 2008; King et al., 2005; Liberti et al., 2008) that has found that land use, land cover, and land slope play an important role in governing erosion.

**Table 5**  
The five interrill–rill erosion intensities and their corresponding descriptive combinations of B<sub>Mj</sub>, B<sub>Js</sub>, LULC and LS.

E <sup>1</sup>	Erosion TYPES <sup>2</sup>	B <sub>Mj</sub> <sup>3</sup>	B <sub>Js</sub> <sup>3</sup>	LULC <sup>4</sup>	LS % <sup>5</sup>
E <sub>S</sub>	S1	D	D	F1, F2	<60
	S1Ri1	D	D	F3	<60
	S1Ri1	D	D	R1, R2	<60
	S1Ri1 + RO <sub>D</sub> <sup>6</sup>	C	C	F2	>40
	S1Ri1 + RO <sub>D</sub> <sup>6</sup>	C	C	R2, R3	>40
E <sub>M</sub>	S2	D	C	R2, R3	0–2
	S2Ri1	D	C	F3	<15
	S2Ri1	D	C	R1, R2, R3	2–5
	S2Ri2	D	C	R1, R2, R3	5–8
	S2Ri2 + RO <sub>A</sub> <sup>6</sup>	D	C	R1, R2, R3	>12
E <sub>H</sub>	S2Ri2 + RO <sub>A</sub> <sup>6</sup>	D	C	F3	>15
	S2Ri3	D	I	R1, R2	8–12
	S3Ri2	D	I	R2, R3	8–25
E <sub>SE</sub>	S3Ri3	D	I	R1, R2, R3	>25
	S3Ri4	C	I	R2, R3	12–25
E <sub>VSE</sub>	S4Ri3	C	I	R2, R3	>15
	S4Ri5	I	I	R3	>25
	S5Ri4	I	I	R2, R3	>15

<sup>1</sup> Interrill–rill erosion intensities (E): slight erosion intensity (E<sub>S</sub>), moderate erosion intensity (E<sub>M</sub>), high erosion intensity (E<sub>H</sub>), severe erosion intensity (E<sub>SE</sub>), and very severe erosion intensity (E<sub>VSE</sub>).

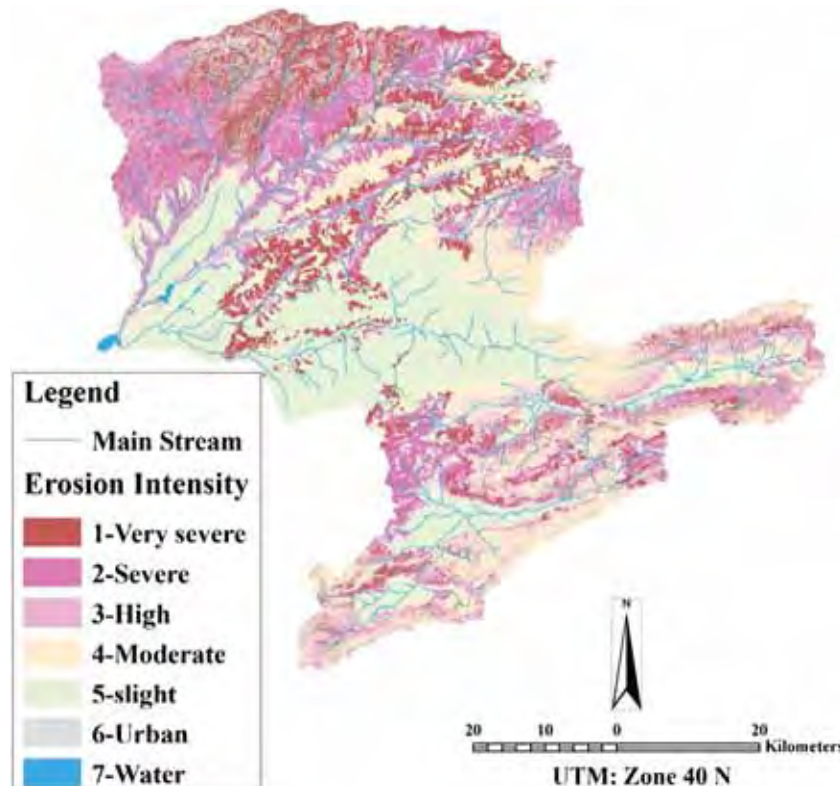
<sup>2</sup> The 15 erosion types as determined by ground-truth data.

<sup>3</sup> RO<sub>D</sub>: surface cover of rock outcrops is more than 80% and RO<sub>A</sub>: surface cover of rock outcrops is 40–80%.

<sup>4</sup> B<sub>Mj</sub>: classes derived by subtracting May (spring) brightness from July (summer) brightness and B<sub>Js</sub>: classes derived by subtracting July brightness from September (late-summer) brightness; with the result that some of these values decreased (D), some increased (I), and the others remained without changing (C).

<sup>5</sup> For land use and land cover (LULC) map ten classes were used (Saadat et al., 2011): irrigated farming (IF), dryland farming (DF); forest (F): high density forest (F1, cover > 70%), medium density forest (F2, 40% < cover ≤ 70%), and low density forest (F3, cover ≤ 40%); rangeland (R): high density rangeland (R1, cover > 30%, mostly between 30% and 50%), medium density rangeland (R2, 15 < cover ≤ 30%), and low density rangeland (R3, cover ≤ 15%); urban (U), and water bodies (W).

<sup>6</sup> LS: land slope.



**Fig. 3.** Soil erosion intensities map.



## 6. Conclusions and recommendations

This study encompassed a relatively large watershed (4511.8 km<sup>2</sup>) which included a variety of erosion types/intensities, from slight to very severe. More than 24% of the land surface experiences severe and very severe erosion. The original contribution of this study is the inclusion of a brightness index ( $B_{MJ}$  and  $B_{JS}$ ) in order to enhance interrill–rill erosion intensity classification. Considering only rangeland and forest (66.2% of the total area), for each interrill–rill erosion intensity it was found that a unique combination of brightness occurred.

The newly proposed approach presented in this paper is expected to be a powerful tool in the creation of interrill–rill erosion intensity maps over large areas, which in turn can be used in the development of soil conservation and watershed management plans, especially in regions with limited high resolution image access and limited accurate data of erosion controlling factors such as vegetation, topography, climate, and soil characteristics. The unique advance made during the development of this protocol is the application of medium resolution satellite images in combination with some ancillary layers to create interrill–rill erosion intensity maps with an accuracy of 96%. It is recommended that further research similar to that presented in this paper be conducted in other regions with differing climatic and geomorphologic characteristics to show whether the proposed protocol presented in this paper can be used effectively in other areas. The analysis as presented in this paper could also be done with satellite images taken at different times of the season. It may be, particularly for other climatic zones, that there is a better time of season for image acquisition that would provide more information. In addition, atmospheric correction is another approach that could be explored in future studies to improve the accuracy of image classifications. This was not explored in this study; however, this is a promising topic for future research.

## Acknowledgments

The authors gratefully acknowledge the Forest, Range and Watershed Management Organization of Iran and the Watershed Management Office of Golestan Province for providing available datasets and facilities, manpower and field support for this study. Partial funding for this study came from a grant held by Jan Adamowski from the Natural Sciences and Engineering Research Council of Canada.

## References

- Alewel, C., Meusburger, K., Brodbeck, M., Banninger, D., 2008. Methods to describe and predict soil erosion in mountain regions. *Landscape Urban Plan.* 88, 46–53.
- Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* 37, 35–46.
- Consulting Engineering, Lar, 2007. The study on flood and debris flow in the Golestan province. Regional Water Board in Golestan. Ministry of Energy, Tehran, Iran (425 pp. (In Farsi)).
- Crist, E.P., Kauth, R.J., 1986. The tasseled cap de-mystified. *Photogramm. Eng. Remote Sens.* 52 (1), 81–86.
- de Vente, J., Poesen, J., 2005. Predicting soil erosion and sediment yield at the basin scale: scale issues and semi-quantitative models. *Earth Sci. Rev.* 71, 95–125.
- de Vente, J., Poesen, J., Verstraeten, G., Van Rompaey, A., Govers, G., 2008. Spatially distributed modelling of soil erosion and sediment yield at regional scales in Spain. *Glob. Planet. Chang.* 60, 393–415.
- Dymond, C.C., Mladenoff, D.J., Radeloff, V.C., 2002. Phenological differences in Tasseled Cap indices improve deciduous forest classification. *Remote Sens. Environ.* 80, 460–472.
- FAO, 2006. Guidelines for Soil Description. Food and Agriculture Organization of the United Nations, Rome (97 pp.).
- Flanagan, D.C., Ascough II, J.C., Nearing, M.A., Lafen, J.M., 2001. The water erosion prediction project (WEPP) model. In: Harmon, R.S., Doe III, W.W. (Eds.), *Landscape Erosion and Evolution Modeling*. Kluwer Academic/Plenum Publishers, New York, pp. 145–199.
- Focardi, S., Loiole, S.A., Mazzuoli, S., Bracchini, L., Dattilo, A.M., Rossi, C., 2008. Satellite-based indices in the analysis of land cover for municipalities in the province of Siena, Italy. *J. Environ. Manag.* 86, 383–389.
- Healey, S.P., Cohen, W.B., Zhiqiang, Y., Krankina, O.N., 2005. Comparison of Tasseled Cap-based Landsat data structures for use in forest disturbance detection. *Remote Sens. Environ.* 97, 301–310.
- Huang, C., Wylie, B., Yang, L., Homer, C., Zylstra, G., 2002. Derivation of a tasseled cap transformation based on Landsat 7 at-satellite reflectance. *Int. J. Remote Sens.* 23 (8), 1741–1748.
- Jain, S.K., Singh, P., Seth, S.M., 2002. Assessment of sedimentation in Bhakra Reservoir in the western Himalayan region using remotely sensed data. *Hydrol. Sci. J.* 47 (2), 203–212.
- Japan International Cooperation Agency (JICA), 2005. The study on flood and debris flow in the Caspian Coastal area focusing on the flood-hit region in Golestan province. Interim Report (Tehran, Iran, 302 pp.).
- Kimaro, D.N., Poesen, J., Msanya, B.M., Deckers, J.A., 2008. Magnitude of soil erosion on the northern slope of the Uluguru Mountains, Tanzania: interrill and rill erosion. *Catena* 75, 38–44.
- King, C., Baghdadi, N., Lecomte, V., Cerdan, O., 2005. The application of remote sensing data to monitoring and modelling of soil erosion. *Catena* 62, 79–93.
- Liberti, M., Simonello, T., Carone, M.T., Coppola, R., D'Emilio, M., Macchiato, M., 2008. Mapping badland areas using LANDSAT TM/ETM satellite imagery and morphological data. *Geomorphology*. <http://dx.doi.org/10.1016/j.geomorph.2008.11.012>.
- Morgan, R.P.C., 1995. Soil Erosion and Conservation. Longman Group Limited, Essex, UK (198 pp.).
- Morgan, R.P.C., 2005. Soil Erosion and Conservation, 3rd edition. Blackwell publishing, Oxford (304 pp.).
- Morgan, R.P.C., Morgan, D.D.V., Finney, J.J., 1984. A predictive model for the assessment of erosion risk. *J. Agric. Eng. Res.* 30 (3), 245–253.
- Parsons, A.J., Wainwright, J., 2006. Depth distribution of interrill overland flow and the formation of rills. *Hydrol. Processes* 20 (7), 1511–1523.
- Poesen, J., Nachtergaele, J., Verstraeten, G., Valentin, C., 2003. Gully erosion and environmental change: importance and research needs. *Catena* 50 (2), 91–133.
- Renard, K.G., Foster, G.R., Weesies, G.A., McCool, D.K., Yoder, D.C., 1997. Predicting soil erosion by water: a guide to conservation planning with the revised universal soil loss equation (RUSLE). *Agriculture Handbook*. vol. 703. US Department of Agriculture, Washington DC (384 pp.).
- Rudra, R.P., Dickinson, W.T., Wall, G.J., 1998. Problems regarding the use of soil erosion models. In: Boardman, J., Favis-Mortlock, D. (Eds.), *Modelling Soil Erosion by Water*. Proceedings of the NATO, Global Environmental Change, vol. 55. Springer, Berlin, pp. 175–189.
- Saadat, H., Bonnell, R., Sharifi, F., Mehuys, G., Namdar, M., Ale-Ebrahim, S., 2008. Landform classification from a digital elevation model and satellite imagery. *Geomorphology* 100, 453–464.
- Saadat, H., Adamowski, J., Bonnell, R., Sharifi, F., Namdar, M., Ale-Ebrahim, S., 2011. Land use and land cover classification over a large area in Iran based on single date analysis of satellite imagery. *ISPRS J. Photogramm. Remote Sens.* 66 (5), 608–619.
- Sharifi, F., Saghafian, B., Telvari, A., 2002. The great 2001 flood in Golestan province, Iran: causes and consequences. Proceedings of the International Conference on Flood Estimation. March 6–8, 2002. Berne, Switzerland, CHR-KHR Report 11–17, pp. 263–271.
- Siakeu, J., Oguchi, T., 2000. Soil erosion analysis and modelling: a review. *Trans. Jpn. Geomorphol.* 21 (4), 413–429.
- Smith, S.J., Williams, J.R., Menzel, R.G., Coleman, G.A., 1984. Prediction of sediment yield from Southern Plains grasslands with the Modified Universal Soil Loss Equation. *Journal of Range Management* 37 (4), 295–297.
- Stehman, S.V., 1999. Basic probability sampling designs for thematic map accuracy assessment. *Int. J. Remote Sens.* 20, 2423–2441.
- Stocking, M.A., Murnaghan, N., 2001. Handbook for the Field Assessment of Land Degradation. Earthscan Publication Ltd., London (169 pp.).
- Vrieling, A., 2006. Satellite remote sensing for water erosion assessment: a review. *Catena* 65, 2–18.
- Vrieling, A., Rodrigues, S.C., Bartholomeus, H., Sterk, G., 2007. Automatic identification of erosion gullies with ASTER imagery in the Brazilian Cerrados. *Int. J. Remote Sens.* 28 (12), 2723–2738.
- Vrieling, A., De Jong, S.M., Sterk, G., Rodrigues, S.C., 2008. Timing of erosion and satellite data: a multi-resolution approach to soil erosion risk mapping. *Int. J. Appl. Earth Obs. Geoinf.* 10, 267–281.
- Wischmeier, W.H., Smith, D.D., 1978. Predicting Rainfall Erosion Losses: a guide to conservation planning. *Agric. Handbook No. 282*. US Department of Agriculture, Washington, DC.



# Technology and Product Reports

## Design, Construction, and Operation of a Demonstration Rainwater Harvesting System for Greenhouse Irrigation at McGill University, Canada

Sadman Islam<sup>1</sup>, Mark Lefsrud<sup>2,3</sup>, Jan Adamowski<sup>2</sup>,  
Blake Bissonnette<sup>1</sup>, and Allison Busgang<sup>1</sup>

**ADDITIONAL INDEX WORDS.** water recycling, water supply, water quality, urban water demand, sustainability, algae growth

**SUMMARY.** Increasing stress on urban water demand has led to the exploration of the potential of rainwater use and water recycling to promote sustainable water resources management. Rainwater harvesting (RWH) not only has the potential to reduce water demand but also contributes to other sustainable objectives, including reducing stormwater pollutant loads, reducing erosion, and inducing natural flow regimes by means of flood control, in urban streams. This research involved the design, construction, and field-testing of an RWH system used to irrigate greenhouses at the Macdonald Campus of McGill University in Quebec, Canada. The purpose of the RWH system was to collect rainwater from a roof area of  $\approx 610 \text{ m}^2$  (the Horticulture Services Building on the Macdonald Campus of McGill University) to meet the irrigation demands of the two Horticulture Research Center greenhouses on the campus ( $\approx 149 \text{ m}^2$  each) from May to October. Over its two years of operation, it was found that the amount of rainwater collected did not only meet the peak irrigation demands of the greenhouses (which amounted to almost 700 gal of water per day), but that there was also enough water for the irrigation of the nearby student-run gardens. The harvested rainwater was clear and did not cause any harm to the plants. The major problem that was experienced during the operation of the RWH system was that of algae growth in one of the water collection tanks. This issue was resolved by covering the tank with metallic green wallpaper, thereby blocking most of the sunlight from entering the tank. The RWH system is currently being used for irrigation and as a demonstration project to promote the learning of sustainable technologies on campus and in the surrounding communities.

We thank the McGill Sustainability Projects Fund for providing us with the required funding for the setup and maintenance of the project.

Department of Bioresource Engineering, McGill University, 21111 Lakeshore Road, Sainte-Anne-de-Bellevue, H9X 3V9, QC, Canada

<sup>1</sup>Research Assistant.

<sup>2</sup>Assistant Professor.

<sup>3</sup>Corresponding author. E-mail: mark.lefsrud@mcgill.ca.

As a part of the family of green technologies, RWH can be defined as the capturing of rain runoff from roofs and other surfaces and storing it for a later purpose (Despins et al., 2009). As an ancient practice, RWH cisterns were common in ancient Greek, Etruscan, Roman, Indian, and other civilizations (Boers

and Asher, 1982). In Jordan for example, surface runoff has been collected for over 4000 years. Elsewhere, archaeological research in Venice resulted in the identification of more than 6000 subterranean rainwater cisterns constructed during the Middle Ages for domestic water supply. More recently, the advent of urban sprawl has resulted in a decrease in the amount of forested lands, wetlands, and other forms of open spaces that absorb and clean storm water in the natural system (Leopold, 1968). This has caused degradation in the water quality of water bodies that are now used in the agricultural and domestic sectors. The practice of RWH has been gaining popularity as the usage of rainwater is much cleaner (in terms of carbon dioxide emissions) than the usage of municipal water supplies. Like other water conservation techniques, RWH is considered to be a viable means to manage urban water resources more efficiently and sustainably (Basinger et al., 2010).

Domestic water usage is a significant component of the global water demand. RWH can be used for both nonpotable and potable purposes such as garden use, toilet flushing, washing clothes, hot water systems, and drinking water supply (Khastagir and Jayasuriya, 2010). Catchment area, storage material, and the distribution system are a few design considerations that have to be taken into account when constructing a RWH system. When selecting a rainwater tank, a house owner often only focuses on the location where the tank will be placed, its aesthetics, and the cost. However, other key design variables that need to be considered are: amount of precipitation in the area, extent of catchment area, and the end use of the water. If the tank is sized properly, the volume of rainwater in the tank will be able to supplement the household water demand; this also reduces the chances of the tank being empty or overflowing. Also, if the catchment area is fairly large, a greater number of end-use applications can allow for greater water savings.

According to the United Nations Environment Programme (UNEP, 2012), examples of RWH and its utilization can be found all across the world. For example, with almost 86% of Singapore's population living in high-rise buildings, using RWH has

become an important component of reducing rising urban water demand in the region. In cities and regions such as Tokyo, Thailand, and China, RWH is seen as a way of mitigating water shortages, controlling floods, and securing water for emergencies. In Bangladesh, where the rural population is plagued by arsenic in the drinking water, RWH is being used to provide a potable water supply, and in St. Thomas, U.S. Virgin Islands it is now mandatory to install a RWH system to acquire a residential building permit.

In Canada, a study in Ontario by Farahbakhsh et al. (2009) indicated that if there are no weather anomalies and the precipitation pattern follows the historical trend, then municipal water demand can be reduced by as much as 47% in domestic households in Ontario, Canada, if RWH is used for domestic water supply. In southeast Brazil, Ghisi (2006) estimated that RWH practices can reduce potable water demand by 48% to 100%. In Germany, where RWH has been in use since the 1980s, potable water demand has decreased by 30% to 60% due to the use of roof runoff harvested in 4- to 6-m<sup>3</sup> tanks, which is then used for toilet flushing (Hermann and Schmida, 1999). RWH is being used in the rural areas of the developing world (such as India, China, and Africa), where it is mainly used for irrigation during dry periods. In such environments, RWH is seen as a measure that can help fight food scarcity (Fox et al., 2005).

The quality of water collected in a RWH system is affected by several factors, which include the proximity to roads and heavy industries, presence of wildlife, and meteorological conditions such as temperature and rainfall patterns in the region (Despins et al., 2009). Although these factors may deteriorate the quality of the collected rainwater, Kumar (2004) states that treatment by prestorage treatment devices such as filtration or first flush diversion can help to expel most of the harmful sediments from the roof (catchment area) into the surroundings and also prevent the stagnation of water in the tanks, thereby restricting the breeding of insects. Similarly, poststorage treatment devices such as ultraviolet disinfection, chlorination, or slow sand filtration can further improve the water quality,

in some cases making the harvested rainwater a source of potable water. As an added precaution, storage tanks must be closed at all times to prevent the entry of insects and to reduce evaporation losses. Finally, the most commonly occurring problem of algal blooms in storage tanks can be mitigated by both chemical and physical techniques (Bartsch, 1954). Chemical techniques include the use of copper sulfate in small concentrations such that the algal bloom is destroyed without harming the roots of the plants (if the water is used for irrigation). Even though the solution is effective, it is not viable as the cost of implementation rises with increasing volumes of water in the storage tank. Covering the storage tanks with a reflective material, which prevents sunlight transmission into the water tanks, is an alternative solution (Bartsch, 1954). This is a significantly less expensive and more practical approach. Another issue that must be addressed in the design of RWH systems is the handling of overflows during large rainfall events. Methods for handling overflow may include onsite infiltration, as well as discharging to an existing storm sewer infrastructure (Farahbakhsh et al., 2009).

The cost of a RWH system varies depending on the size of the catchment area, the type of cisterns chosen to store the rainwater, and the piping material used to transport the water for its end use. Cost was by far the biggest barrier identified for the implementation of RWH systems in a recent study conducted by Farahbakhsh et al. (2009); the typical roof area of a domestic household in Canada is  $\approx 230$  m<sup>2</sup> (Canadian Mortgage and Housing Corp., 2012), which translates to a cost in the range of C\$2000

to C\$8000 (in Canadian dollars) or the setup of a RWH system. The yearly maintenance costs and decommissioning cost at the end of the lifespan of the systems costs an additional C\$2000 (Roebuck and Ashley, 2006). It should be noted that RWH systems have one of the lowest payback periods for sustainable engineering systems (maximum 15 years), and government incentives can often offset some of the expenses (Despins et al., 2009).

The Faculty of Agricultural and Environmental Sciences on the Macdonald Campus of McGill University has been active in the promotion of sustainable technologies. The Horticulture Research Center at the Macdonald Campus serves as a training, education, and research center for students involved in vegetable and fruit production and is used by numerous departments on the campus. The Horticulture Services Building is used as a storage, processing, and retail space for the activities that occur in the surrounding gardens and the two greenhouses. Irrigation of the greenhouses consumes  $\approx 700$  gal of freshwater per day. The objective of the RWH project described in this article is to manage urban water resources more efficiently and promote the learning of practical sustainable technologies at the university and the surrounding communities through educational tours to the project site. Although research on domestic RWH systems has been increasing, there are still relatively few publications that describe the construction and advantages of such systems. This article is intended to fill this gap by explaining cost-effective ways for the setup and maintenance of this sustainable water supply option.

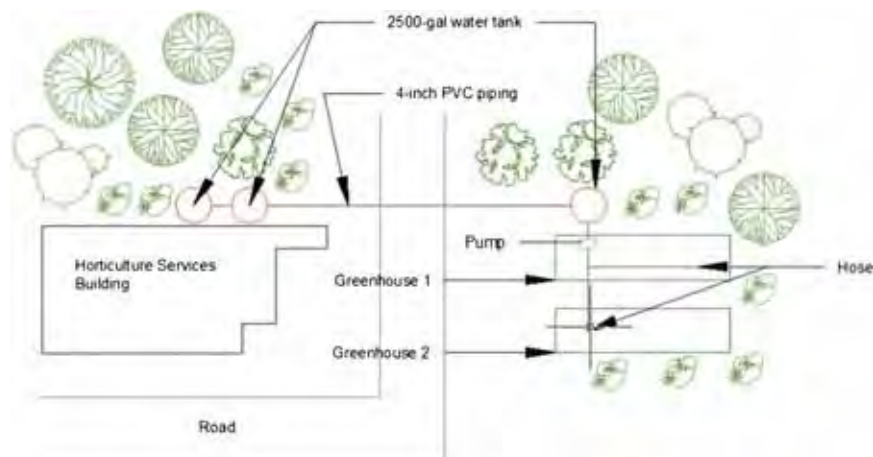
### Units

To convert U.S. to SI, multiply by	U.S. unit	SI unit	To convert SI to U.S., multiply by
29.5735	fl oz	mL	0.0338
0.3048	ft	m	3.2808
0.0929	ft <sup>2</sup>	m <sup>2</sup>	10.7639
0.0283	ft <sup>3</sup>	m <sup>3</sup>	35.3147
3.7854	gal	L	0.2642
0.7457	horsepower	kW	1.3410
2.54	inch(es)	cm	0.3937
25.4	inch(es)	mm	0.0394
1	ppb	$\mu\text{g}\cdot\text{L}^{-1}$	1
1	ppm	$\text{mg}\cdot\text{L}^{-1}$	1
6.8948	psi	kPa	0.1450
$(^{\circ}\text{F} - 32) \div 1.8$	$^{\circ}\text{F}$	$^{\circ}\text{C}$	$(^{\circ}\text{C} \times 1.8) + 32$

## Methodology

The setup of the RWH system (Fig. 1) comprises the Horticulture Services Building and the two Horticulture Research Center greenhouses ( $\approx 50.3$  m away). Rainwater is collected from a sheet metal roof with a catchment area of  $\approx 610$  m<sup>2</sup>. Roof runoff is redirected using 4-inch aluminum eaves troughs into a 2500-gal polyethylene water tank (part no. 40051; Norwesco, St. Bonifacius, MN), with the help of downspouts. An aluminum mesh attached to the top of the eaves troughs prevents leaves and other large debris from entering the water tank. Weather data collected from Environment Canada helped in the sizing of the water tanks by determining the average annual precipitation pattern along with the longest recorded dry period on the island of Montreal. The average annual amount of rainfall in Montreal over the period of 1971–2000 is 834.9 mm, whereas the most prolonged interval without precipitation is 7 d over the past five years (Environment Canada 2012a, 2012b). However, the system was designed to withstand a 1 in 10 year storm event, resulting in the accumulation of 19 m<sup>3</sup> of rainwater over a period of 15 min. It is estimated that the tanks require 52 mm rainfall to fill up with a collective storage volume of 7500 gal, able to sustain the peak irrigation demands of both greenhouses for 11 d.

The collection tanks (Fig. 2) are located next to the Horticulture Services Building. The site is covered with vegetation, and the soil contains a large amount of rocks, adding both aesthetics and stability to the system. The first tank is connected in series to two other identical tanks, each at a decreasing elevation. Gravitational pull and the accumulated head pressure in each tank govern the flow of water between the tanks. The tanks are connected to each other via 4-inch polyvinyl chloride (PVC) pipes, with a system of 4-inch valves and overflows allowing one to control the inflow and outflow of water in each tank. The third tank rests beside the greenhouses (which have an area of  $\approx 149$  m<sup>2</sup> each), and it is connected to a 1-horsepower centrifugal water pump (part no. 4011K11; McMaster-Carr, Santa Fe Springs, CA) that supplies irrigation water to both greenhouses.



**Fig. 1.** Overview map of the rainwater harvesting (RWH) system at the Horticulture Services Building (Macdonald Campus, Sainte-Anne-de-Bellevue, QC, Canada) and at the two Horticulture Research Center greenhouses. Rainwater is collected from a sheet metal roof, directed using 4-inch (10.2 cm) aluminum eaves troughs into a three 2500-gal (9463.5 L) polyethylene water tank.



**Fig. 2.** A side view of the Horticulture Services Building (Macdonald Campus, Sainte-Anne-de-Bellevue, QC, Canada), two 2500-gal (9463.5 L) polyethylene water tanks, and the sheet metal roof directed using 4-inch (10.2 cm) aluminum eaves troughs with downspouts.



**Fig. 3.** A view of the pressure switch at the Horticulture Services Building (Macdonald Campus, Sainte-Anne-de-Bellevue, QC, Canada) that turns the 1-horsepower (0.75 kW) centrifugal water pump on and off for irrigation, when the line pressures are 25 and 40 psi (172.4 and 275.8 kPa), respectively.



With an inlet diameter of 2 inches and an outlet of 1.5 inches, the 1-horsepower centrifugal water pump delivers water at a flow rate of 22 gal/min and 40 psi of pressure at the furthest end of the greenhouses (24 m away). The pump is completely automated with the help of a pressure switch (Fig. 3) that turns the pump on and off when the line pressures are 25 and 40 psi, respectively. The pump helps to meet the peak irrigation needs of the greenhouses from May to October, which can amount up to  $\approx 350$  gal of water per day in each greenhouse. The piping within the greenhouses mainly consists of 1.5-inch semirigid pipes, which allows for flexibility within the structures.

Overflows are designed on all three tanks with 4-inch PVC pipes. The overflows help to discharge the water into the surroundings once the water volume in the tanks reaches 2200 gal. The overflows in the first two tanks are combined together with a “T-junction,” and the water is discharged into the nearby vegetation. The entire system performs efficiently, since it requires no manual labor from the beginning (collection of rainwater) to the end (irrigation of plants).

To verify the chemical and biological composition of the collected rainwater, a thorough water analysis was carried out. Two rainwater samples were collected in 25-mL containers from the first and second storage tanks. The samples were collected 1 week after a rainstorm event (25 Sept. 2012) and stored in ambient temperature ( $\approx 25$  °C). A third sample was collected from the tap that was previously used to supply irrigation water to the two greenhouses, for comparison. The samples were bottled and shipped to a water analysis laboratory that analyzed the samples for various heavy metals as well as the total Kjeldahl nitrogen content using inductively coupled plasma mass spectrometry. Three water samples were collected and stored in the same fashion for the bacterial count. The samples were shipped to the same water analysis laboratory where the total bacterial count was carried out using different plating techniques.

## Results and discussion

This project was the first of its kind to be undertaken at McGill University and serves as a tool for

both irrigation and education. Similar projects can be found at the University of Guelph and McMaster University in Canada, where research is being done to design and operate household RWH systems for flushing toilets, watering gardens, and drinking purposes. To date, the RWH system at the Macdonald Campus of McGill University has been successful throughout the summer and fall in meeting the irrigation demands of both greenhouses combined. Before the RWH system, the town of Sainte-Anne-de-Bellevue, QC, Canada, mainly supplied the irrigation water with some water provided by the Macdonald Campus water supply system (<10%). A decentralized water supply that uses its own source of water is a more sustainable way of providing water for irrigation purposes at the Macdonald Campus. In addition, rainwater is naturally soft water without any minerals, chlorine, fluoride, and other chemicals.

The RWH system not only stores enough rainwater for watering the greenhouses but it has also been able to water the nearby student-run gardens. The design includes steps for collection, filtering, storage, and delivery of the water for irrigation. Although the system did not include steps for the chemical treatment of the stored rainwater, the water that was used for irrigation in the greenhouses was clear and did not have any odor associated with it. Since the

catchment area was a sheet metal roof, it limited the leaching of harmful substances such as lead, mercury, and polycyclic aromatic hydrocarbons into the captured rainwater that would have otherwise been present if the roof was made of asphalt shingles (Van Metre and Mahler, 2003). At the same time, the presence of dissolved and particulate copper is minimized due to the absence of clay or concrete tiles on the roof (Forster, 1996).

Storage tank material can also affect the pH of rainwater that has been collected and change its chemical quality (Despins et al., 2009). The use of high-density polyethylene water tanks in the setup was because plastic tanks create no changes in the pH and chemical composition of the stored water (Hart and White, 2006). Concrete cisterns have reported a rise in the pH from 5 on the roof surface to 9.4 in the tank (Hart and White, 2006). There is also a significant concern about chemical leaching of zinc from metal tanks. In the RWH system at McGill University, sedimentation has played a major role in reducing the contaminant load of the stored rainwater. Sedimentation allows the water to rest in order for the flocculated or coagulated particles to settle out. Ideally the water should be allowed to rest for more than 4 h for the sedimentation process to be effective. The collected rainwater was allowed to rest for 24 h after every

**Table 1. Comparison of levels of heavy metals in rainwater (collected on 25 Sept. 2012, from the roof of the Horticulture Services Building, Macdonald Campus, Sainte-Anne-de-Bellevue, QC, Canada) and a municipal tap water sample (at the same location) to determine water quality of the rainwater harvested for greenhouse irrigation purposes. Samples were collected 1 week after a rainstorm event, stored at ambient temperature [ $\approx 25$  °C (77.0 °F)], and analyzed for various heavy metals using inductively coupled plasma mass spectrometry.**

Parameter	Unit <sup>z</sup>	Reported detection			
		limit	Rain 1 <sup>y</sup>	Rain 2 <sup>y</sup>	Tap 1 <sup>x</sup>
Copper	$\mu\text{g}\cdot\text{L}^{-1}$	3.0	<3.0	<3.0	9.3
Lead	$\mu\text{g}\cdot\text{L}^{-1}$	1.0	<1.0	<1.0	<1.0
Zinc	$\mu\text{g}\cdot\text{L}^{-1}$	3.0	285	359	7.4
Aluminum	$\mu\text{g}\cdot\text{L}^{-1}$	30	<30	<30	45
Iron	$\mu\text{g}\cdot\text{L}^{-1}$	300	<300	<300	<300
Manganese	$\mu\text{g}\cdot\text{L}^{-1}$	5	<5	<5	<5
Sodium	$\mu\text{g}\cdot\text{L}^{-1}$	2000	<2000	<2000	9720
Selenium	$\mu\text{g}\cdot\text{L}^{-1}$	2	<2	<2	<2
Total mercury	$\text{mg}\cdot\text{L}^{-1}$	0.0001	<0.0001	<0.0001	<0.0001

<sup>z</sup>1  $\mu\text{g}\cdot\text{L}^{-1}$  = 1 ppb, 1  $\text{mg}\cdot\text{L}^{-1}$  = 1 ppm.

<sup>y</sup>Two rainwater samples (Rain1 and Rain2) were collected in 25-mL (0.85 fl oz) containers from the first and second storage tanks.

<sup>x</sup>Sample (Tap 1) was collected from the tap that was previously used to supply irrigation water to the two greenhouses, for comparison. The sample was bottled in a 25-mL container.



rainstorm before it was used for irrigation. This provided a source of primary treatment to the collected water.

Water quality also is significantly affected by the type of piping material that is used to transport the water for its end use. PVC pipes were used in this project since they are relatively inert and do not affect the water quality in any significant way. In a comparison between concrete and PVC pipes, Davies et al. (2010) found that concrete drainage systems have a significant influence on water chemistry, particularly where the inflow is acidic. The major factor identified in the research was the presence of calcium, bicarbonate, and potassium ions from the concrete pipes.

As can be seen in Table 1, the chemical composition of the harvested rainwater is fairly similar to tap water. Among the various heavy metal concentrations analyzed, a notable difference can be found between the concentrations of zinc and copper (which are micronutrients for plants). Since the catchment area is made of sheet metal, the average concentration of zinc in the two rainwater samples is  $322 \mu\text{g}\cdot\text{L}^{-1}$ . Although this is  $\approx 43.5$  times higher than the zinc concentration in the tap water sample, the Canadian Council of Ministers of the Environment (CCME, 2011) states that the threshold limit for zinc in irrigation water is  $1000 \mu\text{g}\cdot\text{L}^{-1}$  when the soil pH is below 6.5. Thus, the zinc concentration in the rainwater is not toxic to the greenhouse plants. Copper concentrations however are on average three times lower in the rainwater when compared with tap water. This is can be mainly attributed to the fact that copper pipes are used as a means of delivering the municipal water to the greenhouses, whereas the stored rainwater is transferred to the greenhouses with the help of PVC and plastic piping. Plant deficiency in copper can easily be avoided by fertilization.

Table 2 displays the total nitrogen content and bacterial count for the rainwater and tap water samples. The total nitrogen content in both samples was the same, but the bacterial count in the rainwater was  $\approx 473$  times higher, when compared with the tap water. This indicates that the rainwater cannot be used as a potable water supply, but it does not make

a difference when it is solely used for irrigation.

The tanks in this study were sized to incorporate a suitable safety factor to prevent perpetual overflows and the flooding of the project site. It was calculated that a combined storage volume of 7500 gal would be able to withstand a 1 in 10 year rainstorm event lasting 15 min. Sizing the tanks to have a large storage volume also ensured that the Horticulture Research Center will not deplete its supply of irrigation water during the summer months. However, even with such a considerable storage capacity, some of the water in the tanks had to be expelled (through the overflows) since the tanks did not have enough volume to store the large amounts of water that were sometimes obtained (after big rainstorms). Since the irrigation water demand in the greenhouses declined from September onwards, unused rainwater in the storage tanks accounted for the overflowing of the cisterns. The approximation of the tanks being able to survive a 1 in 10 year rainstorm for 15 min was therefore an underestimate. Addition of another 2500-gal tank (amounting to a total storage volume

to 10,000 gal) should solve the problem of overflows.

There were no problems with the breeding of insects and the entry of foreign objects into the water tanks, as they were sealed with the plastic caps that were provided by the manufacturers. Overall, the cost of the project was  $\approx \text{C}\$8000$  (Table 3). Most of this cost was associated with the installation of the eaves troughs on the roof of the Horticultural Services Building. For ordinary domestic households, this cost would be significantly lower, especially if the eaves troughs are already installed on the roof. For an average household with a roof area of  $230 \text{ m}^2$ , the cost of installation may range between  $\text{C}\$1500$  and  $\text{C}\$2000$  (Save the Rain Campaign, 2010). This cost may change depending on the historic precipitation data of the region and the end use of the harvested rainwater. If new eaves troughs need to be installed, the cost could increase to  $\approx \text{C}\$3000$  to  $\text{C}\$5000$ .

Based on the 2011 annual precipitation data collected from a local weather station,  $411.5 \text{ m}^3$  of rainwater (including overflows) was collected by the RWH system from May to

**Table 2. Comparison of total Kjeldahl nitrogen (TKN) content and the bacterial count (BHAA) in the harvested rainwater (collected on 25 Sept. 2012, from the roof of the Horticulture Services Building, Macdonald Campus, Sainte-Anne-de-Bellevue, QC, Canada) and a municipal tap water sample (at the same location) to determine water quality of the rainwater harvested for the greenhouse irrigation purposes. Samples were collected 1 week after a rainstorm event, stored at ambient temperature [ $\approx 25^\circ\text{C}$  ( $77.0^\circ\text{F}$ )], and analyzed for TKN and BHAA (using dilution plating techniques).**

Parameter	Unit <sup>a</sup>	Rain 1 <sup>b</sup>	Rain 2 <sup>b</sup>	Tap 1 <sup>b</sup>
TKN	$\text{mg}\cdot\text{L}^{-1}$	<1.0	<1.0	<1.0
BHAA	cfu/mL	120,000	230,000	370

<sup>a</sup>1  $\text{mg}\cdot\text{L}^{-1}$  = 1 ppm, 1 cfu/mL = 29.5735 cfu/fl oz.

<sup>b</sup>Two rainwater samples (Rain 1 and Rain 2) and one tap water sample (Tap 1) were collected in 25-mL (0.85 fl oz) containers from the first and second storage tanks along with the tap that previously supplied irrigation water to the two greenhouses.

**Table 3. A detailed cost analysis for the setup of the rainwater harvesting system at the Horticulture Services building located on the Macdonald Campus of McGill University, Sainte-Anne-de-Bellevue, QC, Canada.**

Type of cost	Value (Canadian dollars)
Purchase and installation of eaves troughs	2000
Purchase of 200 ft (61.0 m) of 4-inch (10.2 cm) PVC (PVC) pipes	1000
Purchase of three 2500-gal (9463.5 L) water tanks	3300
Purchase of water pump	300
Purchase of pressure switch	36
Purchase of 100 ft (30.5 m) of semirigid piping	150
Pump and pipe fittings	1200

October. Since the irrigation of the greenhouses required  $\approx 318 \text{ m}^3$  of water during this period of time, all of the irrigation water was supplied by the RWH system. In the city of Montreal, water is priced at a very low rate of C\$0.22/ $\text{m}^3$ . If all of the collected water was used for irrigation, it would lead to a saving of C\$90.53 every year, which results in a payback period of  $\approx 88$  years. Although this is a very long period of time, for water-scarce regions where the price of water is much higher (such as Las Vegas, NV, or the Republic of Cyprus) the payback period will be much faster.

The major problem that was faced in the project was that of algae growth in the final storage tank. The first two tanks were located behind the Horticulture Services Building and were not subjected to large amounts of sunlight throughout the day, reducing problems with algae growth.

However, the final tank was situated in an open space beside the greenhouses, maximizing its exposure to sunlight, particularly during dawn and midday. Although the high-density polyethylene wall of the tank was 1-inch thick, almost 20% of the sunlight was still transmitted to the stored rainwater (MS-01; Apogee Instruments,

North Logan, UT). This created an algal bloom (that was restricted to the surface of the stored water). Before the algae could spread throughout the water body, the tank was flushed and refilled. The tank was then completely wrapped with metallic green wallpaper (as shown in Fig. 4), which helped to reflect almost all the sunlight, preventing any algal blooms. The tank has to be re-wrapped every year due to wear and tear of the wallpaper.

Fresh water scarcity is a major concern around the world, with many regions beginning to lack the water resources necessary to meet the demands of ever growing populations, as well as ecosystems. This project was able to introduce and demonstrate a simple method that can help to address the problem of freshwater scarcity. RWH systems require no treatment of the water (if it is used for irrigation), thereby reducing installation and operational costs. For this project, municipal water for greenhouse irrigation is no longer purchased from the town of Sainte-Anne-de-Bellevue, Quebec, thereby creating a savings of  $\approx$ C\$70 from the month of May to October.

Finally, an additional important outcome was the implementation of a simple and sustainable technology to increase awareness of

RWH on the McGill campus and its surroundings. The project is mainly being used as a demonstration tool to enrich learning on and around the campus on the topic of simple sustainable water technologies. The project helps to show that the setup of a RWH system is relatively simple and is an example of how individual stakeholders can begin to help the transition to more sustainable and decentralized water supply systems. The project allows visitors to appreciate that along with economic incentives (especially in regions where water prices are very high), RWH systems also provide other advantages such as reducing flooding and erosion, as well as decreasing the amount of water that needs to be produced via water treatment plants. Since the completion of the setup, a number of classes from McGill University (as well as outside visitors) have visited the site, including the following courses: BREE-327 (Bioenvironmental Engineering), BREE-510 (Watershed Systems Management), and PLNT-312 (Urban Horticulture).

## Conclusion

Various measures to meet urban water demand and to help communities transition to sustainable water use are currently being explored around the world. For example, rainwater and reclaimed wastewater have been used for a variety of purposes ranging from potable water supply to irrigation and greywater recycling used in toilets and for household cleaning. Rainwater harvesting on a small scale provides an example of one approach to help communities transition to more sustainable water resources management. The effectiveness of RWH systems mainly depend on their geographic locations due to varying amounts of precipitation in different places, the size of the catchment area, and the end use of the harvested rainwater. The Macdonald Campus of McGill University functioned as an ideal study site for such a project. Although the setup of a RWH system in Montreal has very little economic incentives, this may not be the case in water-scarce regions. The system at the Horticulture Research Center was designed to withstand a 1 in 10 year storm event lasting 15 min, to prevent any water shortage that could be



Fig. 4. A 2500-gal (9463.5 L) polyethylene water tank at the Horticulture Services Building (Macdonald Campus, Sainte-Anne-de-Bellevue, QC, Canada) wrapped with metallic green wallpaper to reflect almost all the sunlight, preventing any algal blooms.

faced during irrigation. With a catchment area of 610 m<sup>2</sup>, the RWH system located beside the Horticulture Research Center was able to supply all of the water required for irrigation of the two 149-m<sup>2</sup> greenhouses on the Macdonald Campus of McGill University from the beginning of the summer to the end of the fall. Due to the materials used for the roof (the catchment area), the storage tanks, and the piping, the collected rainwater was clean with very little change in the water chemistry. This made the harvested rainwater ideal for irrigation purposes. The only obstacle that was faced during the operation of the RWH system was algae growth in the final water tank. This issue was resolved by flushing the tank and wrapping it with metallic green wallpaper that reflected most of the sunlight away from the stored water. Not only is the system beneficial for the Horticulture Research Center and the student-run gardens, it also plays an important part in educating the student population at McGill University and the surrounding communities about the benefits of sustainable water resources technologies. Moreover, the RWH system can be expanded in the future to include other buildings located on campus to supply the buildings with their own source of greywater for toilet flushing, floor cleaning, and other purposes. As urbanization continues around the world, RWH provides a useful approach to save water resources.

### Literature cited

Bartsch, A.F. 1954. Practical methods for control of algae and water weeds. *Public Health Rpt.* 68:749–757.

Basinger, M., F. Montalto, and U. Lall. 2010. A rainwater harvesting system reliability model based on nonparametric stochastic rainfall generator. *J. Hydrol. (Amst.)* 392:105–118.

Boers, T.M. and J.B. Asher. 1982. A review of rainwater harvesting. *Agr. Water Mgt.* 5:145–158.

Canadian Council of Ministers of the Environment. 2011. Canadian environmental quality guidelines summary table. 25 Aug. 2012. <<http://st-ts.ccme.ca/?lang=en&factsheet=229>>.

Canadian Mortgage and Housing Corp. 2012. CMHC releases comprehensive report on housing in Canada. 1 Nov. 2012. <<http://www.cmhc-schl.gc.ca/en/corp/nero/nere/2008/2008-11-13-0815.cfm>>.

Davies, P.J., I.A. Wright, O.J. Jonasson, and S.J. Findlay. 2010. Impact of concrete and PVC pipes on urban water chemistry. *Urban Water J.* 7:233–241.

Despins, C., K. Farahbakhsh, and C. Leidl. 2009. Assessment of rainwater quality from rainwater harvesting systems in Ontario, Canada. *J. Water Supply* 58:117–134.

Environment Canada. 2012a. Canadian climate normals 1971–2000. 11 Feb. 2012. <[http://www.climate.weatheroffice.gc.ca/climate\\_normals/results\\_e.html?stnID=5420&lang=c&dCode=0&StationName=MONTREAL&SearchType=Contains&province=ALL&provBut=&month1=0&month2=12](http://www.climate.weatheroffice.gc.ca/climate_normals/results_e.html?stnID=5420&lang=c&dCode=0&StationName=MONTREAL&SearchType=Contains&province=ALL&provBut=&month1=0&month2=12)>.

Environment Canada. 2012b. National climate data and information archive. 26 July 2012. <[http://www.climate.weatheroffice.gc.ca/climateData/dailydata\\_e.html?timeframe=2&Prov=QC&StationID=10873&dlyRange=1993-08-26|2011-07-13&Year=2011&Month=7&Day=01](http://www.climate.weatheroffice.gc.ca/climateData/dailydata_e.html?timeframe=2&Prov=QC&StationID=10873&dlyRange=1993-08-26|2011-07-13&Year=2011&Month=7&Day=01)>.

Farahbakhsh, K., C. Despins, and C. Leidl. 2009. Developing capacity for large-scale rainwater harvesting in Canada. *Water Qual. Res. J. Canada* 44:92–102.

Forster, J. 1996. Patterns of roof runoff contamination and their potential implications on practice and regulation of treatment and local infiltration. *Water Sci. Technol.* 38:39–48.

Fox, P., J. Rockstorm, and J. Barron. 2005. Risk analysis and economic viability of water harvesting for supplemental

irrigation in semi-arid Burkina Faso and Kenya. *Agr. Syst.* 83:231–250.

Ghisi, E. 2006. Potential for potable water savings by using rainwater in the residential sector of Brazil. *Build. Environ.* 41:1544–1550.

Hart, C. and D. White. 2006. Water quality and construction materials in rainwater catchments across Alaska. *J. Environ. Eng. Sci.* 5:S19–S25.

Hermann, T. and U. Schmida. 1999. Rainwater utilization in Germany: Efficiency, dimensioning, hydraulic and environment aspects. *Urban Water* 1: 307–316.

Khastagir, A. and N. Jayasuriya. 2010. Optimal sizing of rainwater tanks for domestic water conservation. *J. Hydrol. (Amst.)* 381:181–188.

Kumar, D.M. 2004. Roof water harvesting for domestic water security: Who gains and who loses? *Water Intl.* 29:43–53.

Leopold, L.B. 1968. Hydrology for urban planning: A guidebook on the hydrologic effects of urban land use. U.S. Geological Survey Circ., U.S. Dept. Interior, Washington, DC.

Roebuck, R.M. and R.M. Ashley. 2006. Predicting the hydraulic and life-cycle cost performance of rainwater harvesting systems using a computer based modelling tool. 7th Intl. Conf. Urban Drainage Modelling. p. 546–553.

Save the Rain Campaign. 2010. Rainwater harvesting FAQ's. 11 Feb. 2012. <<http://www.savetherain.info/media-centre/rainwater-harvesting-faqs.aspx>>.

United Nations Environment Programme. 2012. Examples of rainwater harvesting and utilisation around the world. 11 June 2012. <<http://www.unep.or.jp/ietc/publications/urban/urbanenv-2/9.asp>>.

Van Metre, P.C. and B.J. Mahler. 2003. The contribution of particles washed from rooftops to contaminant loading to urban streams. *Chemosphere* 52:1727–1741.

# Assessing the Potential Impacts of Four Climate Change Scenarios on the Discharge of the Simiyu River, Tanzania Using the SWAT Model

Research Article

Alain Lubini<sup>1</sup> and Jan Adamowski<sup>2,\*</sup>

<sup>1</sup> Department of Civil Engineering, Ecole Polytechnique de Montréal, Canada

<sup>2</sup> Department of Bioresource Engineering, McGill University, Ste. Anne de Bellevue, Quebec, Canada

\* Corresponding author E-mail: jan.adamowski@mcgill.ca

Received 13 Feb 2013; Accepted 26 Mar 2013

DOI: 10.5772/56453

© 2013 Lubini and Adamowski; licensee InTech. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Abstract** The Soil and Water Assessment Tool (SWAT) was used to explore the potential impact of four climate change scenarios on discharge from the Simiyu River in Tanzania, located in the Lake Victoria watershed in Africa. The SWAT model used in this study was calibrated and verified by comparing model output with historic stream flow data for 1973-1976 as well as 1970-1971. SWAT was operated at daily and monthly time steps during both calibration and verification. For the daily-time step verification, the model had a Nash Sutcliffe coefficient of efficiency (NSE) of 0.52 and a correlation coefficient ( $R^2$ ) of 0.72. For the monthly time-step verification, the recorded NSE and  $R^2$  values were 0.66 and 0.70. In developing climate change scenarios within the general patterns defined by the Intergovernmental Panel on Climate Change, predicted increases in CO<sub>2</sub> concentrations were implemented within the constraints of the model's parameterisation by raising, in a seasonally-specific manner, the values of two proxy parameters:

daily baseline temperature and precipitation. Under all scenarios, Simiyu River discharge increased (24-45%), showing the highest increase in the rainy season (March to May), with the greatest increase occurring during the rainy season (March to May). Discharge was influenced to a much greater degree by increases in precipitation rather than by temperature. The increase in river flow predicted by the model suggests that the potential increase in heavy flood damage during the rainy season will increase, which could, in turn, have significant adverse effects on infrastructure, human health, and the environment in the watershed. The SWAT predictions provide an important insight into the magnitude of stream flow changes that might occur in the Simiyu River in Tanzania as a result of future climatic change.

**Keywords** Climate Change, Modelling, River Flow, SWAT, Simiyu River, Africa



## 1. Introduction

From political and social scientists to engineers and hydrologists, the impact of global warming has demanded the attention of research communities around the world as well as of the general public. Climate change is characterized not only by rises in surface temperatures and sea-levels, but also changes in precipitation and decreases in snow cover (IPCC, 2007). Moreover, climatic shifts, as a result of global warming, have been linked to human health problems, water supply shortages, and biodiversity and ecosystem damage. These climatic shifts have already created problems related to human health, water supply shortages, and loss of biodiversity and ecosystem integrity. Many studies conclude that Africa is not only the continent that is the most vulnerable to the impacts of global warming, but also the one that is the least equipped (financially and managerially) to handle them (Perlis, 2009). The majority of Africa's population relies to a large extent on natural resources. This dependency, coupled with fragile governance capacities, could result in severe problems in addressing the challenges of climate change (IISD, 2009). This will necessitate effective and sustainable water resources management and adaptation strategies in the short, medium, and long-term.

In light of water's importance in agriculture, fisheries, and manufacturing, it is important to assess the potential impacts of climate change on the hydrological processes in a watershed. Therefore, the goal of this research was to use the Soil and Water Assessment Tool (SWAT) to measure the likely impacts of four climate change scenarios on the discharge in Tanzania's Simiyu River, located in the Lake Victoria watershed in Africa. Lake Victoria is not only the largest freshwater lake on the continent, but also one of the major sub-basins within the Nile basin. The Simiyu River was selected because of its economic, social, environmental, and political importance in eastern Africa and because few studies have been conducted in this area on this topic. In this study, the SWAT model was used to evaluate fluctuations in seasonal and annual discharges in response to four future climate scenarios in the Simiyu River. The model was calibrated and verified using historical climate and stream flow data in the Simiyu River.

### 1.1 Previous studies of climate change impacts in African watersheds

Apart from Tyson's (1991) early research, relatively little work has emerged on future climate change scenarios focused on Africa. Tyson's climate change scenarios for southern Africa were constructed using results from the first generation of GCM (General Circulation Model) equilibrium  $2\times\text{CO}_2$  experiments. In a further development, Hulme (1994) presented a method for

creating regional climate change scenarios combining GCM results with the newly published IPCC IS92a emission scenarios and demonstrated the application of the method in Africa. In Hulme's study, changes in mean annual temperature and precipitation under an IS92a emission scenario were predicted for the period of 1990 to 2050.

Hulme et al. (1996) took a more selective approach to the use of GCM experiments in describing the 2050 consequences of three future climate change scenarios for the Southern African Development Community (SADC) region of southern Africa. Three different GCM experiments, selected to span the range of precipitation changes predicted by the GCM for the SADC region, allowed for an assessment of some potential impacts and implications of climate change on agriculture, hydrology, health, biodiversity, wildlife and rangelands. However, considerable uncertainty remains regarding the large-scale precipitation changes simulated by GCMs for Africa. Based on such models, Joubert and Hewitson (1997) nevertheless concluded that, in general, precipitation would increase over much of the African continent by the year 2050. As a specific example, their model predicted that by 2050 parts of the Sahel might expect precipitation rates as much as 15% over 1961-90 means.

More recently, Jung et al. (2007) used the meso-scale meteorological model MM5 and the hydrological model WaSim to perform a joint regional climate-hydrology simulation to estimate the effects of anthropogenic influences on the water balance in the Volta Basin, located in West Africa. Their results indicated a decrease in rainfall at the beginning of the rainy season, an increase in rainfall at the height of the rainy season, and a clear increase in temperature.

Lumsden et al. (2009) assessed present, intermediate, and future climate scenarios for Southern Africa to evaluate potential changes in hydrologically relevant statistics of rainfall that could be observed this century as a result of climate change. The scenarios that were accounted for were developed by a downscaling technique simulated by GCMs. According to their study, more rainfall is projected in the east of the region, in the form of more rain days and more days with greater intensity rainfalls. At the same time, the results indicate less projected rainfall along the west coast and the adjacent interior, with the possibility of a slight increase in inter-annual variability.

Using the output data of GCMs in order to address the impacts of climate change in West Africa was studied by Ardoin-Bardin et al. (2009). Using four different models, they were unable to reproduce rainfall volumes in the Sahelian zone, and they had difficulty in simulating the

seasonal dynamics of rainfall in the Guinean zone. Running the HadCM3 (Hadley Centre Coupled Model, version 3) model with new scenarios for future climate conditions, they came to the conclusion that the possible future changes in runoff are highly dependent on rainfall and, hence, on the quality of the output of a given GCM.

Milzow et al. (2011) explored the use of combined satellite radar altimetry, surface soil moisture estimates (SSM) and Gravity Recovery and Climate Experiment (GRACE) total storage change datasets for hydrological model calibration using a SWAT model in the poorly gauged Okavango watershed in Southern Africa. The study found that the use of the combination of the three data sets improved the parametrization of the model. The model showed poor performance at a daily time step due to absence of in-situ precipitation measurements and large variation in the three precipitation datasets, but performed well for long-term scenarios.

In the past decade, studies carried out in the Simiyu River watershed (the study site of the research described in this paper) of the Lake Victoria watershed of Tanzania, were mainly aimed at forecasting stream flow. A Geographic Information System (GIS) based distributed model developed by combining a grid-based water balance model and a flow accumulation/routing model, and operating on a monthly time step at a spatial grid cell resolution of  $0^{\circ}10'$ , performed reasonably well ( $\pm 3\%$  error on an annual basis) in predicting stream flow in this river watershed (Moges, 1998), but fared poorly in simulating the timing and magnitude of hydrographs.

In a more recent study, Mulungu et al. (2007) explored discharge estimation in the Simiyu River watershed using the SWAT model. Using high resolution data, they used the model to estimate the discharge at Ndangalu gauging station (112012 at  $33.568^{\circ}\text{E}$   $2.639^{\circ}\text{S}$ ). Sensitivity and auto-calibration analysis showed that surface model parameters such as the curve number (CN2) and saturated hydraulic conductivity (SOL\_K) were the most sensitive, with a low level of model performance for the verification period.

Rwetabula et al. (2007) used a spatially distributed model (WetSpa) to estimate daily river water discharge in the Simiyu River. Three years of daily observed discharge values measured at the mouth of the river at Lake Victoria were used to calibrate global parameters of the model. Results showed that 38.6% and 61.4% of the total runoff were provided respectively by surface runoff and interflow, while the groundwater flow contribution was nil. The model revealed that the total outflow to Lake Victoria occurs mainly in the wet season, from March to May and from November to January.

In other watersheds in Africa, Githui et al. (2007) used the SWAT model to investigate the impact of climate change on the runoff of the Nzoia River watershed in Kenya. The results showed an acceptable match between the measured and the predicted discharge ( $R^2 > 0.7$ ), with an increase in rainfall, temperature and surface runoff for all the considered scenarios. The study suggested that the 2020s are likely to experience more flooding events as a result of increased rainfall than in the 2050s.

Doktorgrades (2009) studied the impact of climate change in the sub-surface hydrology of the Volta Basin using the MM5 mesoscale climate model and the hydrological WaSiM (Water Flow and Balance Simulation Model) model using the IPCC 2001 climatological data. In his detailed study, Doktorgrades found that the 2007 report of the IPCC lacked new information on African climate change (but highlighted model uncertainties, particularly over tropical Africa). He also found that the inconsistency of different model projections reflected in the low values of the regional climate change index in Giorgi (2006), which relies on regional temperature and precipitation changes in the IPCC multi-model ensemble framework. This model discrepancy clouds the interpretation of the results of uncertain model parameters, which may impact specifically on the simulation of African climate. In light of all of this, IPCC 2001 data was used in Doktorgrades' study to avoid model discrepancy and uncertainty in the interpretation of the model parameters for African climate simulations. For the same reason as the Doktorgrades (2009) study, our study also uses the IPCC 2001 climate data for simulations for studying the climate change impacts on the flow in the Simiyu River watershed.

Most recently, Mango et al. (2011) used the SWAT model to investigate the impacts of land use and climate change in the hydrology of the Mara River Basin in Kenya. Due to the scarcity of climate data in the region, rainfall estimates were derived from the remote sensing data provided by the Famine Early Warning System (FEWS). The NSE values for the observed and simulated flows using the RFE model was 0.43, while the  $R^2$  value was 0.56. Though the correlation results were not sufficient to predict accurate flows in the watershed, the data can be used for better understanding of the hydrological processes in the watershed.

In the research described in this paper, the Soil and Water Assessment Tool (SWAT) was used to explore the potential impacts of four climate change scenarios on discharge from Tanzania's Simiyu River, located in the Lake Victoria watershed. The purpose of the research is to provide insight into the magnitude of stream flow changes that might occur in the Simiyu River watershed as a result of future climatic change, and to assess what

variables affect discharge the most. This information is critical for a better understanding of the potential effects of climate change on the Simiyu River, including potential adverse effects on infrastructure, human health, and the environment. This information is also critical for the development of water resources strategies and policies as well as possible adaptation strategies.

## 2. Methods

### 2.1 The Simiyu watershed

Roughly 5320 km<sup>2</sup> in area, the Simiyu River watershed, located within the Mwanza, Kwimba and Mwasa districts of Tanzania, lies within the confines 33°15'–35°00'E longitude and 2°30'–3°30'S latitude (Figure 1). Arising in the Serengeti Plains at an altitude of about 1882 m AMSL, the river flows in a westerly direction and discharges into Lake Victoria. The water quality of Lake Victoria has been steadily declining due to point and non-point pollution sources from domestic, industrial and agricultural activities. Tanzanian river basins that pollute Lake Victoria are mainly the Mara, Kagera, and Simiyu watersheds (Crul, 1995). The Simiyu catchment is one of the main contributors to the deterioration of Lake Victoria, because it is relatively large, and has many agricultural activities that use agrochemicals (Ningu, 2000), generating high yields of sediments (Lugomela and Machiwa, 2002).

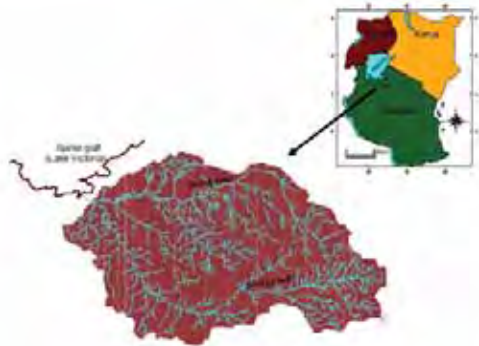


Figure 1. Simiyu Watershed (Rwetabula et.al, 2005)

The watershed is mainly dominated by granite (Syn-orogenic and Magmatite), sedimentary and metamorphic rocks (Nyanzian), with volcanic rocks (Neogene) only covering a small portion of the watershed around the Serengeti plains in north-central Tanzania. Covered by grasses and other herbaceous plants over most of its extent, the watershed's eastern and northern portions are dominated by an open cover of trees, where the crowns do not form a thickly interlaced canopy.

Characterized by a tropical wet and dry climate, the watershed is subject to two wet seasons: the short rains of October through December, and the long rains of March through May. The intermediate months of January and

February represent a period of transition between the two. The dry season occurs between June and September. Mean annual precipitation generally varies between 750 mm and 1100 mm, the latter occurring in the upper part of the watershed. The historical maximum and minimum annual rainfalls are 1500 mm yr<sup>-1</sup> and 400 mm yr<sup>-1</sup>, respectively. Estimates of minimum and maximum mean actual evapotranspiration within the watershed range from 500 to 700 mm yr<sup>-1</sup> and from 1200 to 2700 mm yr<sup>-1</sup>. On average, the monthly discharge varies from 0 to 34 m<sup>3</sup>/s, while during the dry season (June to October), minimum or no discharge is sometimes recorded.

### 2.2 Data

Provided within the framework of the "Friends of the Nile" project, the data required to conduct this study included a digital elevation model, land use, and soil type maps, along with weather data. Daily data records for the climatic variables (precipitation, relative humidity, temperature, wind speed, solar radiation and potential evapotranspiration ( $ET_p$ ), drawn from the two stations located downstream of the Simiyu River within the watershed, were checked for reliability, consistency and homogeneity. The SWAT model used in this study was calibrated and verified by comparing model output with historic stream flow data for 1973-1976 as well as 1970-1971 due to the availability of coexisting flow and climate data during the time period. Given that data series for the time period under concern were incomplete and of variable length, a need clearly existed for either applying some sort of patching technique or calculating an areal mean rainfall. Having chosen the latter approach, daily data from pair meteorological stations were used in a double mass curve analysis over a common period of time in order to assess the homogeneity of the rainfall data. Rainfall distribution was found to be homogeneous within the Simiyu region. However, the resulting distribution of residuals suggested that the outliers be removed in order to calculate the areal mean rainfall using the Thiessen polygon method.

### 2.3 Theoretical background

Developed by the USDA Agricultural Research Service (ARS) to predict the impact of land management practices on water, sediment and agricultural chemical yields over extended periods of time in large complex watersheds (e.g., on a watershed scale) of varying soil, land use and management conditions, the physically based and semi-distributed SWAT model operates on a daily time step (Neitsch et al., 2002). SWAT divides watersheds into sub-watersheds and attendant stream reaches; yields for each sub basin are then computed, routed through stream reaches and reservoirs, and combined as appropriate.

Input information for each sub-basin is grouped or organized into the following categories: climate;

hydrologic response units or HRUs; ponds/wetlands; groundwater; and the main channel, or reach, draining the sub-basin. Hydrologic response units are lumped land areas within the sub-basin that are comprised of unique land cover, soil, and management combinations. An assumption is made that there is no interaction between HRUs within a single sub-basin. Loadings from each HRUs are calculated separately and then summed together to determine the total loadings from the sub-basin (Neitsch et al., 2002).

#### 2.4 Hydrology

The hydrological processes in SWAT are simulated in two phases: land phase and routing phase, the former being based on the water balance equation, expressed in mm of water (Neitsch et al., 2005):

$$SW_t = SW_0 + \sum_{i=0}^{i=n} (R_{day_i} - Q_{surf_i} - E_{a_i} - W_{seep_i} - Q_{gw_i}) \quad (1)$$

where,

$E_{a_i}$  is the amount of evapotranspiration on day  $i$ ,  
 $Q_{gw_i}$  is the amount of return flow on day  $i$ ,  
 $Q_{surf_i}$  is the amount of surface runoff on day  $i$  (mm),  
 $R_{day_i}$  is the rainfall depth on day  $i$ ,  
 $SW_t$  is the final soil moisture content (day  $n$ ),  
 $SW_0$  is the initial soil moisture content (day 0), and  
 $W_{seep_i}$  is the amount of percolation bypass flow exiting the bottom of the soil profile on day  $i$  (mm H<sub>2</sub>O).

To model this phase, SWAT requires climatic variables such as daily precipitation, maximum/minimum air temperature, solar radiation, wind speed and relative humidity.

Surface runoff can be estimated either by the SCS curve number procedure or the Green and Ampt infiltration method. The SCS curve equation is an empirical model developed to provide a consistent basis for estimating the amount of runoff under varying land use and soil types, and is given by:

$$Q_{surf} = \frac{(R_{day} - I_a)^2}{(R_{day} - I_a + S)} \quad (2)$$

where,

$I_a$  represents the initial abstractions – including surface storage, interception and infiltration prior to runoff (mm)  
 $Q_{surf}$  is the accumulated runoff and/or rainfall excess (mm)  
 $R_{day}$  is the rainfall depth for the day (mm), and  
 $S$  is the retention parameter (mm), which varies spatially due to changes in soils, land use, management and slope and temporally due to changes in soil water content. It is related to the curve number (CN) according to the relationship:

$$S = 25.4 \left( \frac{1000}{CN} - 10 \right) \quad (3)$$

The Green and Ampt equation was developed to predict infiltration based on the soil parameters. The method assumes that the soil profile is homogenous and antecedent moisture is uniformly distributed in the profile. The Green-Ampt Mein-Larson infiltration rate is defined as:

$$f_{inf_t} = K_e \left[ 1 + \frac{\Psi_{wf} + \Delta\theta_v}{F_{inf_t}} \right] \quad (4)$$

where,

$f_{inf_t}$  is the infiltration rate at time  $t$  (mm hr<sup>-1</sup>),  
 $F_{inf_t}$  is the cumulative infiltration at time  $t$  (mm),  
 $K_e$  is the effective hydraulic conductivity (mm hr<sup>-1</sup>),  
 $\Psi_{wf}$  is the wetting front matric potential (mm), and  
 $\Delta\theta_v$  is the change in volumetric moisture content across the wetting front (mm mm<sup>-1</sup>)

#### 2.5 Climate change

Climate change simulations though SWAT can be carried out by manipulating the climate inputs such as precipitation, temperature, solar radiation, relative humidity, wind speed,  $ET_p$  and weather generator parameters that are fed into the model (Neitsch et al., 2005). For some parameters (e.g., CO<sub>2</sub> level and temperature) adjustments can be made within individual sub-watersheds by a simple percent increase or decrease in the particular parameter.

Changes in carbon dioxide levels are particularly critical because they have an impact on plant growth. As volume-based atmospheric carbon dioxide concentrations ( $[CO_2]_{vol}$ ) increase, plant productivity increases and plant water requirements go down. Consequently, the canopy resistance term can be modified to reflect the impact of the change in  $[CO_2]_{vol}$  on leaf conductance. Morison and Gifford (1983) found that when the initial  $[CO_2]_{vol}$  ranged between 330 and 660 ppm, a doubling in CO<sub>2</sub> concentration resulted in a 40% reduction in leaf conductance. Consequently, Easterling et al. (1992) proposed the following modification to the leaf conductance term for simulating the effects of  $[CO_2]_{vol}$  on evapotranspiration:

$$g_{l_{CO_2}} = g_l \left( 1.4 - 0.4 \left( \frac{[CO_2]_{vol}}{330} \right) \right) \quad (5)$$

where  $g_{l_{CO_2}}$  is the leaf conductance modified to reflect CO<sub>2</sub> effects (m s<sup>-1</sup>) and CO<sub>2</sub> is the concentration of carbon dioxide in the atmosphere (ppmv).

### 3. Calibration and verification

In order to assess the impact of climate change on discharge from the Simiyu River, the SWAT model had to



first be calibrated and verified. The model was calibrated by comparing predicted and observed hydrographs for the period of January 1973 to December 1976.

The SCS CN method for partitioning of precipitation between surface runoff and infiltration, the variable storage method for channel routing, and the Hargreaves method for calculating  $ET_p$  were model options implemented through the study. These model options were suitable given the level of data availability; for example, the temperature time series was incomplete (maximum and minimum temperatures were only available from 1970 to 1974). Though the Penman Monteith method is more accurate when complete data is available, the Hargreaves method requires less data and can be applied in instances when some weather data are missing (Heuvelmans et al, 2005).

Along with the curve number, the fraction of available water (AW) is an important parameter in calibrating the surface runoff:

$$AW = \frac{\theta_m - \theta_{pwp}}{\theta_{fc} - \theta_{pwp}} \quad (6)$$

where,

$\theta_{fc}$   $\theta_m$  is the soil's volumetric soil moisture content at field capacity

$\theta_{fc}$   $\theta_n$  is the measured volumetric soil moisture content, and

$\theta_{pwp}$   $\theta_{pwp}$  is the soil's volumetric soil moisture content at the permanent wilting point.

Model performance during the calibration and verification phases was evaluated using both the correlation coefficient between measured and predicted parameter values ( $R^2$ ) and the Nash-Sutcliffe coefficient of efficiency ( $NSE$ ). The former measures the strength and direction of a linear relationship between two variables, thus representing the proportion of the variance of one variable that is predictable from the other variable; the latter indicates how well the plot of observed versus predicted data fits a 1:1 line. A perfect match of observed and predicted values would result in an  $NSE$  of 1.0, with values gradually approaching zero as the quality of the match deteriorates to the point ( $NSE < 0.0$ ) where the mean observed parameter is a better predictor than the model (Singh et al., 2004). Values of  $1.0 \geq NSE \geq 0.5$  are generally considered to represent good model performance.

### 3.1 Scenario baseline

A baseline scenario, assumed to reflect current conditions, was executed prior to performing scenario simulations, which were then run with modified climate

inputs for the same simulation period in order to provide a consistent basis for the comparison of scenario impacts. The 4-year period from 1973 to 1976 was used as a baseline condition.

### 3.2 Climate change scenario

Various modeling techniques have been used to develop a wide range of climate change scenarios. Atmosphere-ocean general circulation model simulations (AOGCMs) are often used to help build these scenarios, as are regional climate models that use downscaled AOGCM simulated data. In addition to modeling techniques, scenario-building requires a physical understanding of the processes governing regional responses, as well as a comprehensive understanding of recent historical climate change events. Although the precise pattern of the changes in temperature, precipitation, and extreme events is not well known, simulations by global climate models have led to an agreement on the following trends: i) The global mean surface temperature is projected to increase between 1.1 °C and 6.4 °C by 2100; ii) Global sea levels are projected to rise by 18 to 59 centimeters by the year 2100. Furthermore, climate change scenarios for Africa indicate future warming across the continent, with decadal increases that range from 0.2°C (low scenario) to more than 0.5°C (high scenario) (Hulme et al. 2001; Desanker and Magadza 2001).

In order to examine the sensitivity of stream flow, both altered temperatures and rainfalls were used as a proxy for increases in  $[CO_2]_{vol}$ . To achieve year-long equivalents of rises in  $[CO_2]_{vol}$  of 479, 492, 555 or 559 ppm relative to the baseline, in combination with temperatures being raised over the entire year by 1.5, 2.5, or 4.5°C, rainfall was increased to different degrees on a month-to-month basis (2, 3, 4, 5, 10, 15, or 20%) except in June through August when it remained unchanged (Table 1). The combined increases in temperature and precipitation of the four scenarios implemented were based on putative rises in  $[CO_2]_{vol}$  and IPCC (2001) predictions of future warming across Africa. As discussed earlier, Doktorgrades (2009) found that the IPCC 2007 data for the African continent lacked new information, which was reflected in the inconsistent model projections during simulations. Hence, IPCC 2001 data was used in their study to avoid model discrepancy and uncertainty in the interpretation of the model parameters for African climate simulations. Data from the IPCC Third Assessment Report (2001) was also used throughout the present research project for the same reasons.

Based on the resulting rise in  $CO_2$  the first scenario describes a future world with rapid change in economic structures alongside reductions in material intensity and the introduction of clean and resource efficient technologies. The second scenario is characterized by

intermediate levels of economic development with less rapid and more diverse technological change than in the first scenario. The last two scenarios incorporate the rapid introduction of new and more efficient technologies and place a special emphasis on creating a balanced portfolio of energy sources (i.e., not relying too heavily on one particular source of energy). The third scenario in particular focuses on the use of non-fossil energy sources.

The four scenarios tested in this study were:

**S<sub>492</sub><sup>2.5</sup>** A 2.5°C rise in temperature over the entire year.

Increases in precipitation:  
 Sept-Nov: 3%  
 Dec-Feb: 5%  
 Mar-May: 15%  
 Resulting Rise in [CO<sub>2</sub>]<sub>vol</sub>: 492ppm

**S<sub>479</sub><sup>1.5</sup>** A 1.5°C rise in temperature over the entire year.

Increases in precipitation:  
 Sept-Nov: 1%  
 Dec-Feb: 20%  
 Mar-May: 5%  
 Resulting Rise in [CO<sub>2</sub>]<sub>vol</sub>: 479ppm

**S<sub>555</sub><sup>2.5</sup>** A 2.5°C rise in temperature over the entire year.

Increases in precipitation:  
 Sept-Nov: 4%  
 Dec-Feb: 15%  
 Mar-May: 10%  
 Resulting Rise in [CO<sub>2</sub>]<sub>vol</sub>: 555ppm

**S<sub>559</sub><sup>4.5</sup>** A 4.5°C rise in temperature over the entire year.

Increases in precipitation:  
 Sept-Nov: 2%  
 Dec-Feb: 10%  
 Mar-May: 20%  
 Resulting Rise in [CO<sub>2</sub>]<sub>vol</sub>: 559ppm

Season	S <sub>492</sub> <sup>2.5</sup>			S <sub>555</sub> <sup>2.5</sup>		
	ΔT (°C)	ΔP (%)	ΔCO <sub>2</sub> (ppm)	ΔT (°C)	ΔP (%)	ΔCO <sub>2</sub> (ppm)
Dec.-Feb.	2.5	5	492	1.5	20	479
Mar.-May	2.5	15	492	1.5	5	479
Jun.-Aug.	2.5	0	492	1.5	0	479
Sep.-Nov.	2.5	3	492	1.5	1	479

	S <sub>492</sub> <sup>2.5</sup>			S <sub>559</sub> <sup>4.5</sup>		
	ΔT (°C)	ΔP (%)	ΔCO <sub>2</sub> (ppm)	ΔT (°C)	ΔP (%)	ΔCO <sub>2</sub> (ppm)
2.5	15	555	4.5	10	559	
2.5	10	555	4.5	20	559	
2.5	0	555	4.5	0	559	
2.5	4	555	4.5	2	559	

**Table 1.** Increases relative to the base scenario in air temperature (ΔT), precipitation (ΔP), and atmospheric carbon dioxide concentration (ΔCO<sub>2</sub>) on a volumetric basis, for the four climate change scenarios investigated.

Over the interior, semi-arid regions of the Sahara and central-southern Africa, predicted rises in temperatures ranged from 1.5°C up to 4.5°C (IPCC 2001). Indeed, the intermodal range - an indicator of the extent of agreement between different GCMs - is smallest over North and Equatorial Africa and greatest over the interior of southern Africa. The scenarios described above are drawn from historical data that shows an increase of approximately 0.7°C in temperature across most of Africa during the 20<sup>th</sup> century, a decrease in rainfall over large portions of the Sahelian region, and an increase in rainfall in east central Africa.

Parameter	Description	Units	File type	Soil layer	Adjusted value
SOL_AWC	Soil available water capacity	mm H <sub>2</sub> O mm <sup>-1</sup>	.sol	1	0.12
				2	0.03
SOL_K	Saturated hydraulic conductivity (K <sub>sat</sub> )	mm hr <sup>-1</sup>	.sol	1	0.03
				2	0.50
				1	20
				2	12
1	7				
2	23				
CN2	Curve number		.mgt		40
SLSUBBSIN	Mean slope Length	m			10
SLOPE	Mean slope Steepness	m m <sup>-1</sup>			0.129
OV_N	Manning "n" Value		.hru		0.01
LAT_TIME	Lateral flow Travel	day	.hru		0.001
CH_N 1	"n" value for the tributary		.sub		29.9
CH_K 1	Effective hydraulic conductivity in tributary channel	mm hr <sup>-1</sup>	.sub		11
CH_K 2	Effective hydraulic conductivity in main channel	mm hr <sup>-1</sup>	.Rte		0.019
CH_N 2	Manning "n" value for the main channel		.Rte		0.05
CH_S 2	Average slope of main channel	m m <sup>-1</sup>	.Rte		0.001
GW_DELAY	Ground water Delay	day	.GW		30
ALPHA_BF	Baseflow alpha factor	day	.GW		0.3
GWQMN	Threshold depth of water	mm water	.GW		2575
PND_FR	Fraction of subwatershed area that drains into ponds		.PND		0.09
PND_PSA	Surface area of ponds	ha	.PND		720
PND_K	Hydraulic conductivity through bottom of ponds	mm hr <sup>-1</sup>	.PND		0.005
WET_FR	Fraction of subwatershed that drains into wetlands		.PND		0.05
WET_NSA	Surface area of wetlands at normal water level	ha	.PND		20
WET_MXSA	Surface area of wetlands at maximum water level	ha	.PND		700
WURC	Average daily removal from the reach for the month	m <sup>3</sup> day <sup>-1</sup>	.wus		5×10 <sup>4</sup> 200×10 <sup>4</sup>

**Table 2.** SWAT model parameters after calibration and verification.

Future changes in mean seasonal rainfall in Africa are less well defined. Under the lowest warming scenario, up to 2050 few areas would experience changes in December through February, or June through August precipitation, exceeding two standard deviations of natural variability. However, in parts of equatorial East Africa, rainfall is predicted to increase by 5-20% from December through February, and decrease by 5-10%, from June through August (IPCC 2001b).

#### 4. Results and discussions

In order to create an accurate simulation of the Simiyu River's flow using SWAT, it was necessary to adjust several model parameters to best match observed and simulated values (Table 2). Following this calibration, specific model parameters were set and remained unchanged for the verification period. The comparison between the predicted and observed flows was carried out for both daily and monthly time steps (Table 3).

Statistics*	Daily		Monthly	
	Calibration	Verification	Calibration	Verification
<i>n</i>	1461	730	48	24
Var $Q_{obs}$	403.66	1004.7	178755	447067
SSQ	278374	355262	2584101	3659731
$R^2$	0.73	0.72	0.83	0.83
NSE	0.53	0.52	0.70	0.66

\**n*, number of values; Var  $Q_{obs}$ , variance of the observed flow; SSQ, sum of square deviations between observed and predicted flow values;  $R^2$ , linear correlation coefficient between observed and predicted flow values; NSE, Nash-Sutcliffe coefficient of model efficiency.

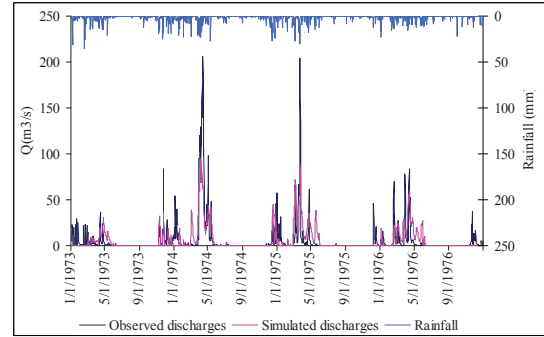
**Table 3.** SWAT model accuracy statistics for calibration and verification phases for both daily and monthly time steps.

#### 4.1 Daily Time step

##### 4.1.1 Calibration

In general, the model was able to accurately simulate discharge rates below 80 m<sup>3</sup>/s, but it failed to simulate higher peaks. This may be partly attributable to inconsistencies in the data set, since some peaks are not explained by the magnitude of the precipitation recorded for the corresponding period. Furthermore, given that the river watershed is mainly dominated by sandy loam soil (which has a high infiltration rate even when thoroughly wetted), these results suggest that ground water storage may be abstracting water from either runoff or from the river itself.

A plot of predicted and measured daily stream flows (Figure 2) over the 4-year calibration period (1973-1976) shows that the model underestimated the flow for the years 1973 and 1974 by 18% and 23%, respectively, and overestimated it by 15% and 21% in 1975 and 1976.



**Figure 2.** Sequential plot of the daily observed and modeled discharges and the rainfall at Ndagalu for the calibration period 1973-1976

The  $R^2$  was 0.73, and the NSE was 0.53, for the calibration period as shown in Table 5.

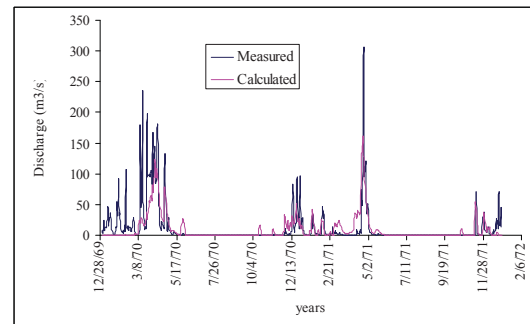
Statistics	Daily time step	
	Calibration	Verification
<i>N</i>	1461	730
Var $Q_{obs}$	403.06	1004.68
SSQ	278374.3	355261.8
<b>R</b>	<b>0.7314</b>	<b>0.7232</b>
<b>NSE</b>	<b>0.5272</b>	<b>0.5156</b>

*n*, number of values; Var  $Q_{obs}$ , variance of the observed flow; SSQ, sum of square deviations between observed and predicted flow values;  $R^2$ , linear correlation coefficient between observed and predicted flow values; NSE, Nash-Sutcliffe coefficient of model efficiency.

**Table 5.** Statistics for daily data

##### 4.1.2 Verification

The verification period encompassed the hydrological years of 1970 and 1971; during this two-year simulation period, no changes were made to the calibrated parameters. The results show that the discharges predicted from the model were not very different from those observed during the calibration period (Figure 3). The calculated flow is less than what was measured for both years, the differences being 37% and 20% for 1970 and 1971, respectively. The  $R^2$  was 0.72, and the NSE was 0.52, for the verification period as shown in Table 5.



**Figure 3.** Predicted and simulated stream flow for the verification period 1970-1971

## 4.2 Monthly time step

### 4.2.1 Calibration

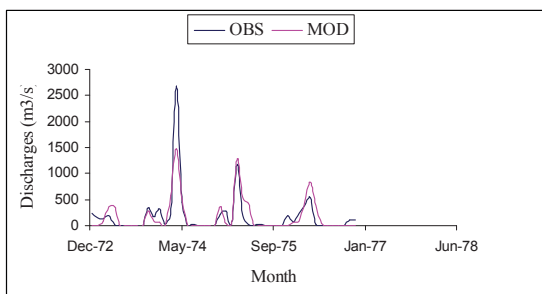
The monthly stream flow data that was collected during the four-year calibration period (1973-1976) revealed that predicted flows closely followed observed flows (Figure 4). The  $R^2$  was 0.83, and the NSE was 0.70, for the calibration period as shown in Table 6.

Statistics	Monthly time step	
	Calibration	Verification
N	48	24
VarQobs	178754.96	447066.8
SSQ	2584100.52	3659731
R	0.8346	0.8273
NSE	0.6988	0.6589

$n$ , number of values; Var  $Q_{obs}$ , variance of the observed flow; SSQ, sum of square deviations between observed and predicted flow values;  $R^2$ , linear correlation coefficient between observed and predicted flow values; NSE, Nash-Sutcliffe coefficient of model efficiency.

**Table 6.** Statistics for monthly data

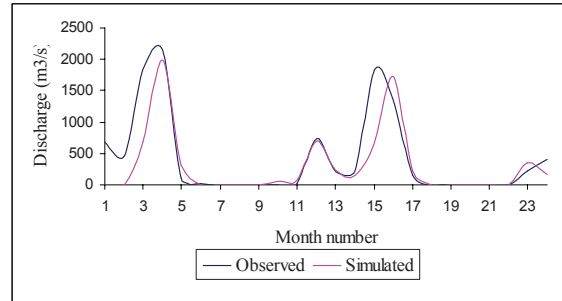
The time-series of predicted and monthly measured stream flows that were used to produce these results show that there was a slight over prediction during high flow months (i.e., March-May) in 1975 and a large under estimation during the same monthly period in 1974; this is most likely because of the uncertainties in the measured values. The data showed a remarkable difference in the magnitude of extreme events; for example, the recorded stream flow during March 1975 was 1176.2  $m^3/s$ , while stream flows in March 1974 measured only 319.79  $m^3/s$ .



**Figure 4.** Monthly observed and calculated discharge for the calibration period January 1973 - December 1976

### 4.2.2. Verification

When analyzing the performance on a monthly time-scale, the verification period shows that the model performed well, although it slightly under predicted the flow for the wet season. However, there was little difference between predicted and measured values for November and December 1970 or for January 1971 (Figure 5).



**Figure 5.** Monthly observed and calculated discharge for the verification period 1970/01 - 1970/12

Respective  $R^2$  and NSE values of 0.83 and 0.66 (Table 6) confirm this conclusion as they are similar to corresponding values for the calibration period ( $R^2=0.83$  and  $NSE=0.70$ ). These verification results indicate that the SWAT model was able to accurately replicate monthly cumulative stream flow levels at the outlet of the watershed over the time period simulated.

In order to develop, verify, and improve the accuracy of the model, a sensitivity analysis was used to identify the factors that had the greatest impact on predicted river flow (see Appendix). These factors were found to be: i) Water capacity of the soil layer (SOL\_AWC); ii) Mean slope length (SLSUBBSN); iii) Mean slope steepness (SLOPE); and iv) Threshold depth of water (GWQMN).

Month	Scenario				
	Baseline	S <sub>492</sub> <sup>2.5</sup>	S <sub>479</sub> <sup>1.5</sup>	S <sub>555</sub> <sup>2.5</sup>	S <sub>559</sub> <sup>4.5</sup>
Jan	1.32	1.67	2.86	2.44	2.05
Feb	0.97	1.28	2.54	2.09	1.67
Mar	19.40	22.87	21.59	22.35	24.27
Apr	26.84	32.45	29.24	30.95	34.40
May	12.15	14.90	13.38	14.25	16.25
Jun	0.27	0.42	0.31	0.36	0.48
Jul	0.01	0.01	0.01	0.14	2.39
Aug	0.00	0.26	0.67	0.76	0.80
Sep	0.00	0.00	0.00	0.00	0.87
Oct	0.00	0.93	0.84	1.49	2.90
Nov	2.29	3.14	3.99	4.37	4.26
Dec	3.68	6.03	7.71	7.22	6.62
<b>Yearly cumulative</b>	<b>66.95</b>	<b>83.96</b>	<b>83.14</b>	<b>86.42</b>	<b>96.95</b>

**Table 4.** Predicted relative changes in river flows ( $m^3 \text{ day}^{-1}$ ) for the 4 climate change and baseline scenarios.

In addition to the curve number, the fraction of available water (AW) is an important parameter in calibrating the surface runoff. For sandy loam soil, predicted river flow was very sensitive to AW; a 0.1 increase in the latter resulted in a 0.31-0.53 increase of the NSE that indicated a better fit, while  $R^2$  remained quite stable. However, a similar change in AW for clay loam soil did not significantly affect the performance criteria. The optimal values of  $\theta_m$  for sandy loam and clay loam soils were 0.6 and 0.5 respectively, both greater than the respective soil's  $\theta_{fc}$ , indicating that one



may expect significant percolation from the top layers to the underlying layers.

Comparatively, the following factors each had a lesser effect on overall model performance: the curve number (CN), base flow alpha factor (ALPHA\_BF), ground water delay (GW\_DELAY), saturated hydraulic conductivity (SOL\_K), groundwater “revap” coefficient, and REVAPMN (threshold depth of water in the shallow aquifer for revap). For example, despite the hydrological response unit (HRU) chosen within the watershed, an increase in curve number from 35 to 79 resulted in neither the NSE nor the R<sup>2</sup> being drastically affected; with the NSE dropping from 0.53 to 0.51 and R<sup>2</sup> decreasing by 1%. An optimal CN value of 40 was obtained for soils allowing moderate water transmission and infiltration when thoroughly wetted (U.S Natural Resource Conservation Service Soil Hydrologic Group B).

#### 4.3 Climate change

Cumulative monthly stream flows (Table 4) predicted under the four scenarios were 24% ( $S_{479}^{1.5}$ ) to 45% ( $S_{559}^{4.5}$ ) greater than those under the baseline scenario. The greatest increases occurred in the rainy season (March to May). Scenarios  $S_{492}^{2.5}$  and  $S_{555}^{2.5}$  had identical rises in temperature, but  $S_{555}^{2.5}$  represented a greater equivalent rise in [CO<sub>2</sub>] than  $S_{492}^{2.5}$ , while  $S_{555}^{2.5}$  represented an overall greater increase in precipitation than  $S_{492}^{2.5}$ . Apart from the long rainy period, the predicted streamflow was greater under  $S_{555}^{2.5}$  than  $S_{492}^{2.5}$ , indicating that the discharge was strongly influenced by the increase in precipitation. For all four scenarios, the magnitude of the discharges recorded during the dry period (June through August) increased slightly compared to the baseline despite the fact that the precipitations were held at 0 mm. An illustration of that increase occurred under  $S_{559}^{4.5}$ , where seasonal discharge [July-August] was 13-fold greater than the corresponding baseline. Under such circumstances, water shortage problems would be alleviated with possible benefits for agricultural activities.

This result is in accordance with the conclusions drawn from the study conducted by Githui et al. (2008) in western Kenya, which is in the same region as the Simiyu River watershed. They found that streamflow response was not sensitive to shifts in temperature with 6-115% change (in streamflow) when rainfall varied from 2.4-23.2%.

The predicted rise in river flow (Figure 6) indicates that the potential for heavy flood damage during the rainy season of March through May is possible. A comparison between the baseline and scenarios 1 and 4 suggests that there will be an increase of 17% and 22% of stream flow during this period. Stream flow is high as a result of an increase in rainfall; this implies high runoff due to the soil

composition, which is mainly made of clay-loam (73.05%) and sand clay loam (25.73%) (Mulungu et al. 2007). The increase in flow as shown above will require adjustment from policy makers in order to address the risk related to more rainfall, with measures needing to be taken to protect cities against likely flooding and waterborne diseases (such as malaria, cholera, and dysentery).

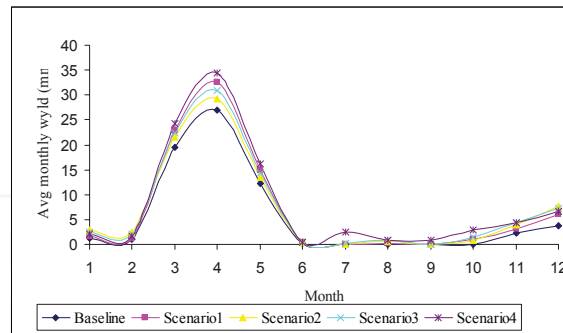


Figure 6. Change in average monthly stream flows predicted for scenarios 1 - 4 relative to the baseline over the 4 year simulation period

#### 5. Conclusions and recommendations

Because of the strategic importance of the Simiyu River in relation to Africa’s vulnerability to climate change, this study chose to analyze the impact of four possible future climate change scenarios on the discharge of the Simiyu River watershed, as measured at the Ndangalu gauging station. This study used a SWAT model that was based on a four year calibration period and a two year verification period, with both calibration and verification being performed on a daily and monthly basis. The simulated climatic scenarios used in this study were hypothetical in nature, and thus cannot be viewed as assessments of absolute future climatic conditions. Nevertheless, the SWAT predictions derived from this study provide insight into the potential magnitude of stream flow changes that could occur as a result of future climatic changes, and the calculated discharges may serve as a comparative baseline for various other climate change scenarios.

In this study, the most sensitive parameters identified were the SOL\_AWC, SLSUBBSN, SLOPE, and GWQMN, and notwithstanding apparent inconsistencies in the data series used, the model was fairly accurate in predicting low flows, but did poorly in predicting extreme peak flow events. This is likely due to high annual and variability, which creates higher variance in the simulated streamflow. Even though sensitivity analyses on individual parameters were conducted, it would also be advisable to perform advanced uncertainty analysis in the future to decreasing the influence of parameter uncertainty on the modeling results. Both approaches should be considered during

watershed delineation. Future studies using the SWAT model should also be done with longer time series in order to improve the output models for the watershed.

For each scenario, increases in precipitation had a greater impact on the river's discharge than increases in temperature, and this was found to be true on both a seasonal basis as well as over longer periods of time. This supports the findings of Wigley and Jones (1984), who evaluated the influences of precipitation changes and direct [CO<sub>2</sub>] effects on stream flow and found that for most rivers in the world annual stream flow was more sensitive to precipitation changes than to the rate of evapotranspiration. The forecasted trends for the long rainy season that were produced during the different simulation scenarios indicated that under increased precipitation, flooding would intensify. The consequences of severe flood damage could negatively impact communication infrastructure, farms, human health, human settlements, and biodiversity. In some areas, commercial ranching may marginally improve as a result of increased rainfall, whereas communal ranching might be exposed to greater risk because of increased erosion and the incursion of woody weeds.

Predicting future rainfall patterns is extremely difficult and this created uncertainties in this study. This is due to the fact that the models used in this study did not account for the role of the El Niño southern oscillation (ENSO) or other atmospheric processes affected by land cover change and dust loadings. Such limitations make it difficult to quantify some of the impacts of climate change, which is why future research should focus on integrated approaches, as well as investigating the links between climate, hydrological, and ecosystem models. Furthermore, because climate change has very regional impacts, climate change scenarios generated by other climate models should be compared with scenarios set up by the GCMs because it will increase our understanding of potential impacts on water resources in Africa in the 21<sup>st</sup> century. There is a specific need to develop regional models that incorporate local climate aspects (from either sparse observations or low resolution numerical simulations) that can be used in constructing various climate change scenarios. Research should also focus on evaluating strategies to sustain and improve the development of the Simiyu River and its watershed in a changing environment. Strategies that incorporate climate change adaptation options will also enhance this process.

## 6. Acknowledgments

The advice of Prof. Willy Bauwens of Vrije University is appreciated. The financial assistance of the Flemish Interuniversity Council, the IUPWARE program, the Katholieke Universiteit Leuven, and an NSERC Discovery Grant held by Jan Adamowski are acknowledged.

## 7. References

- [1] Ardoin-Bardin S., Dezetter A., Servat, E., Paturol J.E., Mahe G., Niel H., Dieulin C., (2009), Using general circulation model outputs to assess impacts of climate change on runoff for large hydrological catchments in West Africa, *Hydrological Sciences* 54(1), 77-89
- [2] Crul, R.C.M., (1995), *Limnology and hydrology of Lake Victoria*. Verhandlungen Internationale Vereinigung Limnologie, 25, 39-48.
- [3] Desanker P.V., Magadza C, (2001), Africa. In: McCarthy, J.J., O.F. Canziani, N.A. Leary, D.J. Doken and K.S. White (eds.). *Climate Change 2001: Impacts, Adaptation and Vulnerability*. IPCC Working Group II, Third Assessment Report, Cambridge University Press.
- [4] Doktorgrades D.E., (2009), *Modelling impacts of climate change on water resources in the Volta Basin, West Africa*, Diese Dissertation ist auf dem Hochschulschriften server der ULB Bonn. [http://hss.ulb.uni-bonn.de/diss\\_online\\_elektronisch\\_publiziert](http://hss.ulb.uni-bonn.de/diss_online_elektronisch_publiziert)
- [5] Easterling W.E., Rosenberg N.J., McKenney M.S., Allan Jones C., Dyke P.T., Williams J.R., (1992), Preparing the erosion productivity impact calculator (EPIC) model to simulate crop response to climate change and the direct effects of CO<sub>2</sub>, *Agricultural and Forest Meteorology*, 59 (1992), pp. 17-34.
- [6] Ficklin D. L., Luo Y., Luedeling E., Zhang M., (2009), Climate change sensitivity assessment of a highly agricultural watershed using SWAT, *Journal of Hydrology*, Volume 374, Issues 1-2, 30, Pages 16-29, ISSN 0022-1694.
- [7] Githui F., Gitau W., Mutua F., Bauwens W., (2009), Climate change impact on SWAT simulated streamflow in western Kenya, *International Journal of Climatology*, 29, 1823-1834.
- [8] Heuvelmans G., Garcia-Quijano J.F., Muys B., Feyen J., Coppin P., (2005), Modeling the water balance with SWAT as part of the land use impact evaluation in a life cycle study of CO<sub>2</sub> emission reduction scenarios, *Hydrological Processes* 19, 729-748
- [9] Houghton J.T., Ding Y., Griggs D.J., Noguera M., Van der Linden P.J., Dai X., Maskell K., Johnson C.A., (2001), *Climate Change 2001: The Scientific Basis*. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. IPCC (Intergovernmental Panel on Climate Change), Cambridge University Press, Cambridge, New York
- [10] Hulme M., (1994) *Regional climate change scenarios based on IPCC emissions projections with some illustrations for Africa area*, Blackwell Publishing on behalf of The Royal Geographical Society (with the Institute of British Geographers) 26(1): 33-34

- [11] Hulme M., (1996), Climate change and southern Africa: an exploration of some potential impacts and implications in the SADC region, CRU/WWF, Norwich, UK, 104pp
- [12] Hulme M., Doherty R., Ngara T., New M., Lister D., (2001), African climate change: 1900–2100, *Climate Research*, 17, 145–168
- [13] IISD, (2009), Climate change and conflict: Lessons from community conservancies in northern Kenya, Conservation development centre, International Institute of sustainable development and saferworld.
- [14] IPCC, (2001), Climate Change. The Physical Science Basis. Cambridge, UK: Cambridge University Press. 996 p.
- [15] IPCC, (2007), Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- [16] Joubert A.M., Hewitson B.C., (1997), Simulating present and future climates of southern Africa using general circulation models. *Prog Phys Geogr* 21:51–78
- [17] Jung G., Kunstmann H., (2007), High-resolution regional climate modeling for the Volta region of West Africa, *Journal of Geophysics Research*, 112
- [18] Lugomela C., Machiwa J., (2002), Nutrient distribution, phytoplankton abundance and species composition in Magu bay following the principal rain season in 2001, in: *Nutrients, microalgae sedimentation and sediment associations at the mouth of Simiyu river (Magu bay of Speke gulf) Lake Victoria, Tanzania, Lake Victoria Environmental Management Project (LVEMP), Tanzania*, 22–39,
- [19] Lumsden T.G., Schulze R.E., Hewitson B.C., (2009), Evaluation of potential changes in hydrologically relevant statistics of rainfall in Southern Africa under conditions of climate change, *Water SA* 35(5), 649–656
- [20] Morison J.I.L., Gifford R.M., (1983), Stomatal sensitivity to carbon dioxide and humidity, *Plant Physiology*, 71: 789–796.
- [21] Mulungu D.M.M., Munishi S.E., (2007), Simiyu River catchment parameterization using SWAT model, *Physics and Chemistry of the Earth* 32(15-18): 1032-1039
- [22] Mango L. M., Melesse A. M., McClain M. E., Gann D., Setegn S. G., (2011), Land use and climate change impacts on the hydrology of the upper Mara River Basin, Kenya: results of a modeling study to support better resource management, *Hydrology and Earth System Sciences*, 15, 2245–2258.
- [23] Milzow, C., Krogh P.E., Bauer-Gottwein P., (2011), Combining satellite radar altimetry, SAR surface soil moisture and GRACE total storage changes for hydrological model calibration in a large poorly gauged catchment, *Hydrology and Earth System Sciences*, Vol.15, 1729-1743.
- [24] Moges, S., (1998), GIS-based distributed water balance model for Lake Victoria basin. University of Dar es Salaam Publication. 124p.
- [25] Neitsch S.L., Arnold J.G., Kiniry J.R., Williams J.R., King K.W., (2002), Soil and Water Assessment Tool, Theoretical documentation. Version 2000, Texas Water Resources Institute, College Station, TWRI Report TR-192.
- [26] Neitsch S.L., Arnold, J.G., Kiniry J.R., Williams J.R., King, K.W., (2005), Soil and Water Assessment Tool Theoretical Documentation. Version 2005, College Station, TX: Texas Water Resource Institute.
- [27] Ningu J., (2000), An inventory of agro-chemicals in the Lake Victoria basin, Mwanza, Tanzania, Lake Victoria Environmental Management Project (LVEMP), Tanzania, 1–23.
- [28] Rwetabula J., Desmedt F., Rebhun M., (2007), Prediction of runoff and discharge in the Simiyu River (tributary of Lake Victoria, Tanzania) using the wetspa model, *Hydrology and Earth System Sciences Discussions*, 4(2), 881-908.
- [29] Singh J., Knapp H. V., Misganaw D., (2004), Hydrologic Modeling of the Iroquois River Watershed using HSPF and SWAT, Watershed Science Section, Illinois State Waters Survey.
- [30] Tyson P.D., (1991), Climatic change in Southern Africa: past and present conditions and possible future scenarios climatic change, *Global Change: Issues for the Southern Hemisphere*, 18, 241-258

# Towards adaptive and integrated management paradigms to meet the challenges of water governance

J. Halbe, C. Pahl-Wostl, J. Sendzimir and J. Adamowski

## ABSTRACT

Integrated Water Resource Management (IWRM) aims at finding practical and sustainable solutions to water resource issues. Research and practice have shown that innovative methods and tools are not sufficient to implement IWRM – the concept needs to also be integrated in prevailing management paradigms and institutions. Water governance science addresses this human dimension by focusing on the analysis of regulatory processes that influence the behavior of actors in water management systems. This paper proposes a new methodology for the integrated analysis of water resources management and governance systems in order to elicit and analyze case-specific management paradigms. It builds on the Management and Transition Framework (MTF) that allows for the examination of structures and processes underlying water management and governance. The new methodology presented in this paper combines participatory modeling and analysis of the governance system by using the MTF to investigate case-specific management paradigms. The linking of participatory modeling and research on complex management and governance systems allows for the transfer of knowledge between scientific, policy, engineering and local communities. In this way, the proposed methodology facilitates assessment and implementation of transformation processes towards IWRM that require also the adoption of adaptive management principles. A case study on flood management in the Tisza River Basin in Hungary is provided to illustrate the application of the proposed methodology.

**Key words** | IWRM, management and transition framework (MTF), management paradigms, participatory modeling, water governance, water management

## INTRODUCTION

The persistence of water issues in many parts of the world has given rise to innovative concepts that advocate an integrated approach to address complexity and uncertainty. Integrated Water Resource Management (IWRM) is the most prominent of such concepts that stresses the importance of integrated and participatory management processes and reform of water governance systems (Medema *et al.* 2008). The term ‘management’ refers to operational activities including the operation, monitoring, strategic planning and implementation of measures, whereas the term ‘governance’ comprises the rules under which a management system operates and different actors and networks help develop and implement water policies (cf., Pahl-Wostl 2009). However, theories and methods for sustainable water resource management and governance are still in the developmental phase and continuous

experiments in application are required to determine effective approaches for research and practice (cf., Galaz 2007; Medema *et al.* 2008). Methodologies are needed that deal with real-world complexity in order to find effective solution strategies, and facilitate knowledge transfer between science, policy, engineering and local communities. Even though IWRM is a concept widely aspired to, it is often still rooted in a traditional ‘predict and control paradigm’ despite its linkage to the idea of adaptive management (cf., Jeffrey & Gearey 2006). The need for the integration of adaptive approaches like policy experimentation and learning into IWRM is increasingly acknowledged in order to effectively realize the concept of IWRM (Galaz 2007).

Management paradigms are appropriate concepts to systematically and comprehensively analyze the interlinkages between a resource system (e.g., groundwater resources),

**J. Halbe** (corresponding author)

**C. Pahl-Wostl**

Institute of Environmental Systems Research,  
University of Osnabrück,  
Barbarastr. 12,  
49076 Osnabrück,  
Germany  
E-mail: jhalbe@uos.de

**J. Sendzimir**

International Institute for Applied Systems Analysis  
(IIASA),  
Schlossplatz 1,  
A-2361 Laxenburg,  
Austria

**J. Adamowski**

Department of Bioscience Engineering,  
McGill University,  
21 111 Lakeshore,  
H9X 3V9,  
Ste Anne de Bellevue,  
Quebec,  
Canada



water management system (e.g., infrastructure) and water governance system (i.e., regulatory structures and processes). A specific analysis of the multiplicity of elements of water resource management paradigms has been discussed by Pahl-Wostl *et al.* (2011). They define a management paradigm as ‘a set of basic assumptions about the nature of the system to be managed, the goals of managing the system and the ways in which these goals can be achieved’ (Pahl-Wostl *et al.* 2011). By being explicit about underlying paradigms, inconsistencies in water management and governance systems become apparent (as discussed by Jeffrey & Gearey (2006)). For instance, public participation can be applied in a ‘predict and control water management paradigm’ (i.e., stakeholders are only informed and consulted), as well as in a ‘community involvement paradigm’ (i.e., co-management of stakeholders). The ignorance of underlying paradigms can lead to miscommunication and subsequent management problems (e.g., stakeholders expect active involvement but can become frustrated due to missing opportunities for engagement).

This paper presents a methodology for the participatory analysis of management and governance systems that supports the design and implementation of transformation processes towards sustainable water management. In this methodology, participatory model building is applied to elicit case specific water management paradigms held by stakeholders. This information is then used to comprehensively analyze the management and governance system through the application of the Management and Transition Framework (MTF) developed by Pahl-Wostl *et al.* (2007, 2010). Based on such an analysis of the status quo, the methodology allows for the participatory envisioning and design of pathways towards sustainable water management and governance. In this way, the proposed methodology facilitates the development, assessment and implementation of strategies towards sustainable water resources governance and management.

The methodology builds upon the management paradigm concept developed by Pahl-Wostl *et al.* (2011) and the use of group model building for the analysis of paradigms (cf., Sendzimir *et al.* 2007). Innovative elements of the methodology proposed in this paper are the conceptualization of sub-system and overall-system paradigms, and the delineation of a structured action research process including elicitation, analysis and assessment of paradigms. Another innovative element of the proposed methodology is the application of the MTF for the design of pathways towards implementation of integrated and adaptive water management. The approach presented in this paper goes beyond

theoretical explanations of institutional and policy change (cf., Cashman 2009) by allowing for participatory analysis and active governance of transformation processes.

This paper is structured in two parts. First, the underlying concepts and methods upon which the proposed methodology is built are presented. Following this, a case study on flood management in the Tisza River Basin in Hungary is provided to illustrate the application of the proposed methodology.

## METHODS

The proposed methodology combines an analysis of the overall water management and governance system (using the MTF) with an investigation of embedded management paradigms applying the systems thinking method in a participatory modeling process. Participatory modeling supports the analysis of case-specific elements of resource management and governance systems, while the MTF analysis allows for a broader perspective by integrating the detected elements found via participatory modeling into an overall system perspective.

### Participatory model building

Different participatory modeling approaches exist that follow different objectives and apply a range of methods (cf., Jonsson *et al.* 2007; Renger *et al.* 2008; Voinov & Bousquet 2010). In this paper, the method of participatory model building using systems thinking and system dynamics methods is proposed for analyzing the perceptions of stakeholders on the management and governance system. Systems thinking is a method for the qualitative analysis of systems and their dynamic behavior through time. System dynamics modeling is based on this qualitative analysis and comprises the quantitative simulation of systems to discover their inherent dynamics as well as allowing for the testing of strategies.

Causal Loop Diagrams (CLDs) are powerful tools of the systems thinking method for the qualitative analysis of systems. They help to depict the system’s structure, and mark time delays and feedback processes that are often responsible for difficulties in controlling the inherent dynamics of the management system. In these diagrams, elements of the system are connected by arrows having positive and negative polarities. A positive link indicates the parallel behavior of variables: in the case of an increase in the ‘cause’ variable, the variable that is affected also increases, while a decrease

in the ‘cause’ variable implies a decrease in the affected one. A negative link indicates an inverse linkage between variables (see Figure 1 as an example of a CLD).

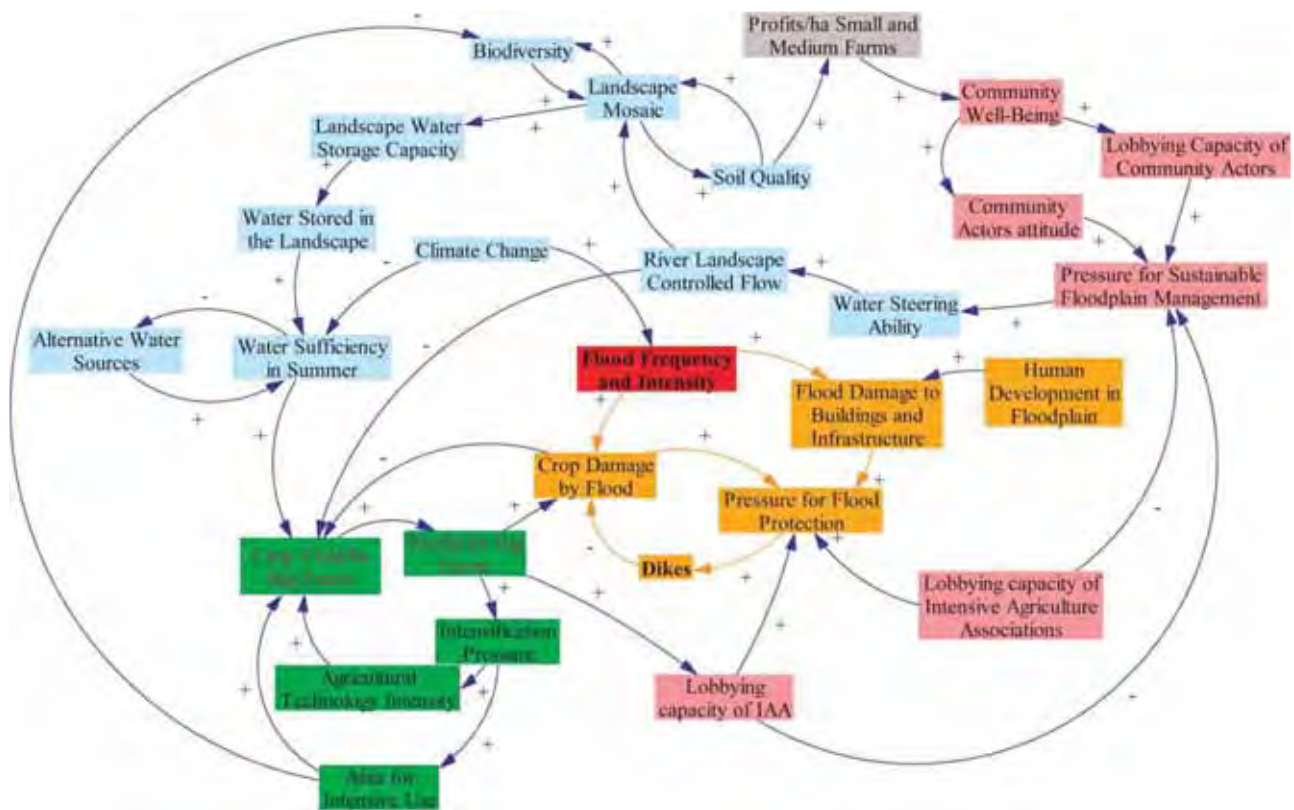
Despite the fact that expert models often offer comprehensive and scientifically validated results, missing ownership and understanding of the model by decision-makers and other stakeholders often impedes the implementation of model-based recommendations. This has led to the development of system dynamics applications that involve stakeholders in quantitative model building. Group model building processes (also called mediated modeling) for collaborative management of complex human-environment problem situations have begun to be applied more frequently over the course of the last decade (e.g., Costanza & Ruth 1998; van den Belt 2004; Tidwell et al. 2004; Metcalf et al. 2010; Halbe et al. accepted).

### Management and transition framework (MTF)

Understanding processes of change towards sustainable resource management and governance requires an analytical approach that allows for the analysis of the

interdependence between structural context and process characteristics. The MTF developed by Pahl-Wostl et al. (2010) supports such analyses of water governance regimes and transition processes towards more adaptive and sustainable systems. The MTF builds upon the three conceptual pillars of adaptive management (cf., Holling 1978), social learning and transformation processes (cf., Pahl-Wostl et al. 2007), as well as the Institutional Analysis and Development Framework (which is aimed at the analysis of the role of institutions in collective choice processes, cf., Ostrom (2005)). Specific emphasis is given to the analysis of adaptive capacity and multi-level learning processes. However, the MTF is not constrained to one specific theory; instead it provides a flexible language that can be tailored to specific research questions (for examples of applications of the MTF, see Schlüter et al. (2010); Sendzimir et al. (2010)).

The MTF helps to formalize structural elements of a water system (which are denoted as ‘classes’) as well as policy and learning processes (cf., Pahl-Wostl et al. 2010). Central classes in the MTF are as follows. An ‘Action Situation’ refers to formal or informal social



**Figure 1** | Causal Loop Diagram regarding the flooding problem in the Tisza River Basin (extended from Sendzimir et al. (2007)). The colors reflect the sub-system paradigms: ‘Adapt to Floods’ paradigm (blue); ‘Control Floods’ paradigm (orange); ‘Economies of Scale’ paradigm (green); ‘Tradition’ paradigm (grey); ‘Community Involvement’ paradigm (pink). The problem variable is highlighted in red. Please refer to the online version of this paper to see this figure in color: <http://www.iwaponline.com/wst/toc.htm>.

processes that lead to relevant outcomes for water management. Results can be ‘Institutions’ (e.g., a new water legislation), ‘Knowledge’ (e.g., increased understanding of stakeholder problem perspectives) or ‘Operational Outcomes’ (i.e., direct physical interventions in the system such as the implementation of infrastructure or distribution of water to different uses). The ‘Action Arena’ class sets the context for the management of a specific water-related problem such as flood management, and is characterized by ‘Strategic Management Goals’, ‘Actors’ and a number of ‘Action Situations’. In this way, the MTF provides a common language to analyze and discuss complex management and governance systems in research and practice. Relational databases are used to support formalization and standardization of data collection and analysis protocols (cf., [Knieper \*et al.\* 2010](#)). A graphical interface allows for the straightforward presentation and discussion of analyses ([http://www.yworks.com/en/products\\_yed\\_about.html](http://www.yworks.com/en/products_yed_about.html)).

A management process can be depicted as a temporal sequence of action situations that are linked by institutions, knowledge or operational outcomes (see example in [Figure 3](#)) and represent different phases in an overall policy cycle (e.g., policy formulation or implementation). Another approach to analyze a management and governance system is the interpretation of action situations as governance and management functions (e.g., water purification or allocation or conflict resolution). While temporal analysis is more suitable to examine the evolution of management issues over time, functional analysis allows for the comprehensive analysis of the status quo of management and governance systems at specific points in time.

## RESULTS AND DISCUSSION

In this section, the proposed methodology is described in more detail and its application is illustrated with an example of flood management in the Tisza Basin, Hungary. The Tisza is a transboundary river and extends from the Ukrainian Carpathian mountains along the Romanian border, flows across the great Hungarian plain and enters the Danube in the Serbian Republic. It is the largest tributary of the Danube with a total catchment area of 157,200 km<sup>2</sup>. In the Hungarian reach of the Tisza Basin, a centralized water management regime has existed since the 19th century, with a focus on engineered flood protection through the large-scale construction of dikes in order to allow for

intensive agriculture and to protect residential and industrial areas. Rising flood intensities and frequencies have resulted in significant challenges for the existing water management paradigm over the past decade. A bottom-up learning process was formed by activists and academics that brought innovative ideas into the flood policy debate. However, a transition towards alternative paradigms has stalled due to weak linkages between the informal learning process and the formal institutions ([Sendzimir \*et al.\* 2010](#)). To address issues such as this, the methodology proposed in this paper supports a structured action research process of water management and governance systems, comprising three steps: (1) elicitation of management paradigms at the sub-system level; (2) analysis of the status quo in management and governance regimes; (3) design of pathways to overcome detected barriers towards sustainable water management.

## Methodology

A management paradigm is defined by a specific ‘system perspective’ regarding the management problem, chosen ‘solution strategies’, as well as ‘risk and uncertainty management strategies’ (cf., [Pahl-Wostl \*et al.\* 2011](#)). The proposed methodology differentiates between paradigms linked to the sub-system level (e.g., the social, environmental, or technical system) and the overall system level (i.e., comprising the complete management and governance system). The former are called ‘sub-system paradigms’ and the latter ‘overall-system paradigms’.

Management paradigms can co-exist at the sub-system level, either by being complementary (i.e., reinforcing each other) or competing (i.e., balancing each other). The proposed methodology builds on the notion that a concerted set of paradigms is usually needed, each tailored to the specific sub-system, to find effective and sustainable solutions. For instance, technical sub-systems (e.g., infrastructure) can be managed by a ‘control paradigm’ that aims at controlling the behavior of the sub-system (different paradigms are presented in detail below, cf., [Table 1](#)). Selected social issues (e.g., an allocation system) can be governed by a ‘community paradigm’ that builds upon the self-organization capacity of stakeholders (e.g., installing irrigation associations). Several paradigms can also belong to the same sub-system by complementing each other. For instance, water pricing can be implemented by applying a ‘market paradigm’ (i.e., prices are set by demand and supply) as well as a ‘control paradigm’ (i.e., the range of prices is pre-determined). Alternative sub-system paradigms can also

**Table 1** | Management paradigms elicited in Figure 1

Dimension	Name				
	'Economies of Scale' Paradigm	'Control Floods' Paradigm	'Adapt to Floods' Paradigm	'Community Involvement' Paradigm	'Tradition' Paradigm
System Perspective	Big farms	River and protected values	Floodplain landscape	Flood prone communities	Small farms
Solution Strategies	Economies of scale; rationalization	Build dikes	River-landscape controlled flows	Community involvement	Traditional farming methods
Risk and Uncertainty Management	Reduce flooding risk and uncertainties	Reduction of uncertainty	Accept flood risk; adaptive management (through experimentation)	Uncertainty dialogue	Build on experience from the past

co-exist by being linked to different locations. For example, an 'adapt to floods paradigm' can be applied in rural areas where retention areas are available, while a 'control floods paradigm' might be more likely to be implemented in urban areas due to fewer adaptation options.

An encompassing 'overall-system paradigm' is linked to the overall resource system and can emerge from the sub-system level (e.g., through the supersession of other sub-system paradigms), or can be purposefully implemented by a higher-level institution (e.g., a ministry for water). In the case of a 'control paradigm' at the overall system level, heterogeneity of sub-system paradigms is constrained as only a limited number of paradigms are compatible with this overall-system paradigm. However, an 'integrated and adaptive overall-system paradigm' (cf., Pahl-Wostl et al. 2011) allows for the coordination of various sub-system paradigms and increases the adaptive capacity of the overall management system.

The proposed methodology that allows for the elicitation of management paradigms at the sub-system level (Step 1) and overall system level (Step 2), as well as the visioning of pathways towards sustainable water management (Step 3) is presented in the following sections.

### Step 1: 'Elicitation of sub-system specific management paradigms'

Participatory model building using systems thinking can support the elicitation of sub-system specific management paradigms from individual participants or groups. The interviewee/group is asked to include the causes and consequences of the particular problem (i.e., the 'system perspective'), as well as preferred 'solution strategies' (e.g.,

technical approaches like building dams or socio-economic aspects like stakeholder involvement). The resulting CLDs will comprise elements of the resource system (e.g., variables like 'precipitation' or 'vegetation type'), the management system (e.g., 'dams' or 'retention areas'), and the governance system (e.g., 'public participation' or 'water legislation'). Further information about the 'risk or uncertainty management' strategies is needed to derive management paradigms from CLDs. Uncertainties are commonly typified in ontological and epistemological uncertainties. While the former denotes complex phenomena whose behavior cannot be predicted, the latter refers to incomplete knowledge or information about a system that can be attained through scientific research (Walker et al. 2003). Relational uncertainties are a third type of uncertainty and acknowledge subjective perceptions of actors. People perceive objects differently depending on personal values, roles, and interests (Brugnach et al. 2008). Based on this categorization, possible strategies for the handling of uncertainties and risks comprise the following (cf., Brugnach et al. 2008):

- Acceptance of uncertainty since *ontological uncertainties* imply that predictions cannot be made.
- Reduction of uncertainty since *epistemological uncertainties* can be minimized through purposeful research.
- Uncertainty dialogue since relational uncertainties requires a dialogue between stakeholders. For instance, the method of participatory model building facilitates the learning of groups and the revision of mental models and frames of participants. The goal of this reframing process is not to determine a 'true' frame – rather, the process aims at widening individual frames to one that considers multiple values and interests of stakeholders.



Figure 1 shows a CLD that was developed via a group model building processes in the Tisza River Basin (see Sendzimir *et al.* 2007). The CLD contains case-specific elements of the resource (e.g., soil quality), management (e.g., dikes), and governance system (e.g., lobbying capacity of community actors).

The group model in Figure 1 integrates different aspects of the flooding problem in the Tisza Basin that were mentioned by stakeholders including government representatives, local activists, and scientists. The CLD shows two different paradigms that are directly related to the flooding problem (represented by the problem variable 'Flood Frequency and Intensity'), namely the 'Adapt to Floods' paradigm (related elements are marked in blue in Figure 1) and the 'Control Floods' paradigm (marked in orange). For the specific attributes of paradigms, see Table 1. Further paradigms are indirectly related to flooding, including the 'Economies of Scale' paradigm (marked in green) of the intensive agriculture sector, and the 'Tradition' paradigm (marked in grey) that is more applicable to small farms. In addition, there are societal processes that demand more of a 'Community Involvement' paradigm (marked in pink) that relies on dialogue between actors.

The group model presents the interplay of different paradigms held by stakeholders related to the flooding problem in the Hungarian reaches of the Tisza Basin. The model clarifies the specific system elements that are related to paradigms. Thus, the CLD can support a purposeful discussion and handling of management paradigms.

## Step 2: 'Analysis of management paradigms embedded in the overall management and governance system'

Management paradigms cannot only belong to the sub-system level but also belong to the overall system. Management paradigms of the overall system can be considered as a general 'mindset' that dominates in the management and governance system. Such overall-system paradigms are not only represented in the way of governing water resources, but are also manifested in infrastructure, information management, and finance, amongst others (cf., Pahl-Wostl *et al.* 2011).

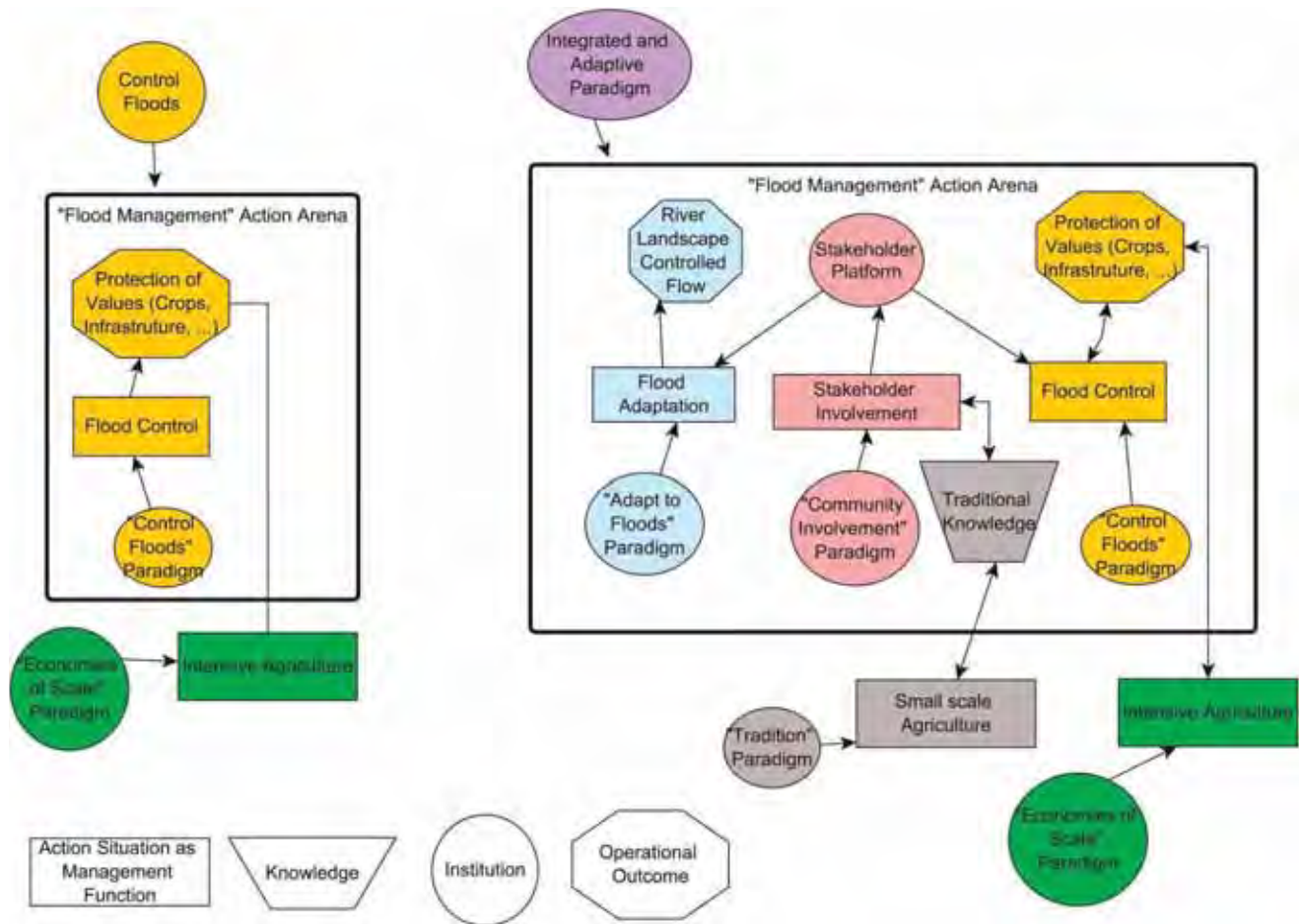
Overall-system paradigms can influence paradigms at the sub-system level. For instance, a 'Control' overall-system paradigm might only be compatible with the 'Economies of Scale' sub-system paradigm due to a similar uncertainty and risk strategy (cf., Figure 1 and Table 1) but usually hampers the functioning of other paradigms at the sub-system level. However, an overall 'Adaptive

and Integrated' paradigm, as conceived by Pahl-Wostl *et al.* (2011), allows for the functioning of all paradigms at the sub-system level due to its integrative nature. This permits diversity in management approaches, which increases the resilience of the social-ecological system (cf., Folke 2006).

Management paradigms pertain to both structural as well as process characteristics of management and governance systems. In the MTF, sub-system paradigms are considered as a type of cultural-cognitive institution and can therefore be an input or output of action situations (i.e., sub-system paradigms are social rules that can influence water management processes). Each action situation is characterized by at least one paradigm. The designation of paradigms to an action situation depends, first, on participating stakeholders who include their perspectives, and second, on input factors (e.g., knowledge, institutions, operational outcomes).

Overall-system management paradigms have a more general influence on the water system and are therefore linked to the Action Arena. As mentioned above, an 'Integrated and Adaptive' paradigm can function as such an overall-system paradigm. However, a 'Predict and Control' paradigm might be the most prevalent current overall-system paradigm, which hampers the application of most of the paradigms at the sub-system level (and thereby limits diversity). Figure 2 shows a graphical representation of the management and governance system in the Tisza Basin comprising action situations as management functions (boxes), institutions (circles), knowledge (trapezes), and operational outcomes (octagons).

On the left hand side of Figure 2, the 'status quo' of the management and governance regime in the Hungarian Tisza Basin is depicted. It is characterized by the dominance of a 'predict and control' paradigm at the overall-system and sub-system level that hampers solution strategies other than 'flood protection'. Intensive agriculture is tightly linked to this kind of management and governance system (Sendzimir *et al.* 2007). On the right hand side, a future vision of an alternative system is presented, which is based on the combination of paradigms included in the group model (see Figure 1). In this future system, an 'integrated and adaptive' flood management paradigm has emerged at the overall-system level, which allows for a diversity of paradigms at the sub-system level (cf., Grabs *et al.* 2006). A 'stakeholder platform' coordinates the 'Flood Control' and 'Flood Adaptation' management functions. Small-scale and intensive agriculture are exogenous elements that are linked to this Action Arena.

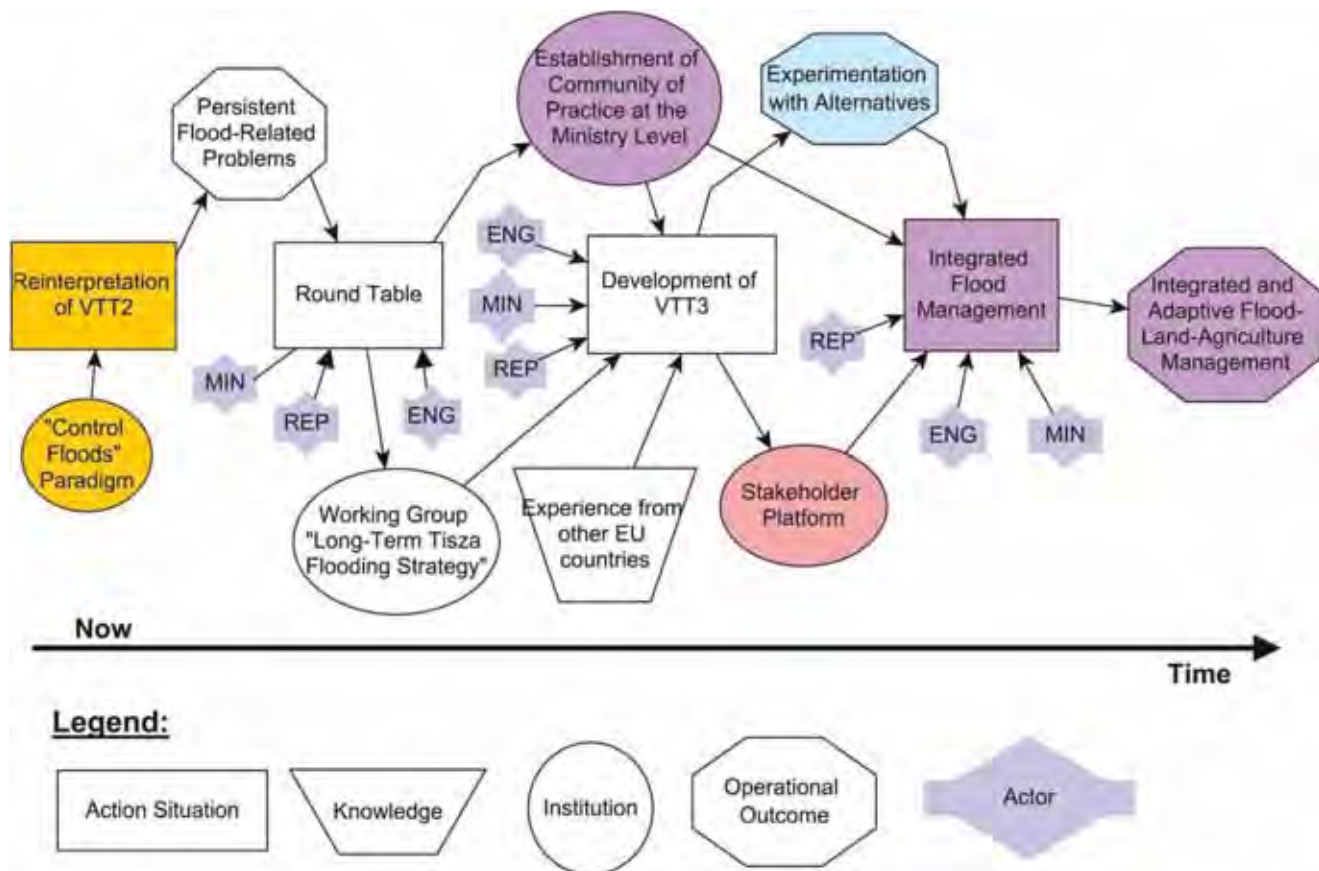


**Figure 2** | Functional analysis of the management and governance system (purple colored variables reflect the new 'integrated and adaptive paradigm'; other colors refer to sub-system paradigms, cf., Figure 1). Please refer to the online version of this paper to see this figure in color: <http://www.iwaponline.com/wst/toc.htm>.

### Step 3: 'Visioning of pathways towards sustainable water management'

Based on the functional diagnosis of the status quo and the desired future of the management and governance system (cf., Step 2), necessary measures through time are defined in Step 3 to achieve the desired changes. The transformation towards sustainable water management and governance usually requires substantial investments of resources and reform of institutions. The MTF can be applied to explore these required changes in a holistic manner, and specify a temporal succession of action situations in order to achieve the desired future. The outcome from this task is the design of a concrete pathway towards a sustainable water management and governance system. The resulting pathways include requirements for institutional change, dissemination or production of knowledge, physical interventions (i.e., 'operational outcomes'), as well as participation of actors.

Figure 3 shows an example pathway towards an integrated and adaptive flood management system for the Hungarian reaches of the Tisza River Basin that was developed based on literature review and expert opinion. The pathway starts from a present action situation comprising the reinterpretation of a flood management policy called VTT2 (in Hungarian: 'Vásárhelyi TervTovábbfejlesztése') (cf., Sendzimir *et al.* 2010), which reflects the mindset of the dominating 'Control Floods' paradigm. The transformation process ends with an 'Integrated and Adaptive Flood Management' action situation that delivers the outcome of 'integrated and adaptive flood-land-agriculture management' (see Figure 3). In the example pathway depicted in Figure 3, two action situations are envisioned to lead to the desired outcome. First, a round table is advised under participation of the Ministry of Environment and Water, responsible engineers and water managers, and representatives of the 'Living Tisza Alliance', which is formed of local activists,



**Figure 3** | Example of a possible pathway towards a sustainable water management and governance system in the Tisza River Basin (purple colored variables reflect the new 'integrated and adaptive paradigm'; other colors refer to the respective management paradigm, cf., Figure 1). Abbreviations of actors mean the following: REP = representatives of 'Living Tisza Alliance'; ENG = Responsible Engineers and Water Managers; MIN = Ministry of Environment and Water. Please refer to the online version of this paper to see this figure in color: <http://www.iwaponline.com/wst/toc.htm>.

national non-governmental organizations and the Village Municipalities Association (cf., Sendzimir et al. 2007). This round table sets the rules for a working group for a long-term flood strategy (institution) that includes all participating actors, and facilitates the institutionalization of a community of practice at the ministry level. The community of practice is supposed to link different departments and facilitate a continuous deliberative process on innovative solutions for the flood management problems in the Tisza Basin. In the past, this was temporarily achieved by the leadership of an individual at the ministry level, before his departure from the parliamentary committee stopped this process. A community of practice could lead to a more sustainable network that is less vulnerable to change of personnel. These activities together with experience from other EU countries with innovative flood management could set the basis for the development of a new flood management policy (named VTT3) that could lead to the institutionalization of a stakeholder platform at the local

and regional scales as well as more experimentation with alternative approaches (e.g., through pilot studies). The established institutions and the experimental approach would result in the 'Integrated and Adaptive Flood Management' action situation that brings together the ministerial 'community of practice' and the stakeholder platform. This leads to river-landscape controlled flows (operational outcome), which reflects a transformation of the overall management and governance system towards a more 'integrated and adaptive' overall-system paradigm.

The example application of the proposed methodology demonstrates a structured approach to analyze water management and governance systems in an integrated way, and envision transformation processes towards IWRM. Through the use of the proposed methodology, hidden assumptions about the management and governance system, as well as requirements for integrated and adaptive solution strategies, are elicited. This approach can be particularly useful in transboundary river basins as paradigms



might vary considerably between international water authorities due to different historical and cultural developments. While the Tisza case was confined to the Hungarian reach of the Tisza, a basin-wide study would be very useful to reveal differences in management paradigms in order to explore opportunities to deal with them constructively. The proposed methodology builds on straightforward methods and is therefore particularly suitable to be applied in collaborative management processes. At the same time, the proposed methodology composed of participatory modeling and the MTF allows for in-depth analysis of the water system. In this way, the linkage of the MTF and participatory modeling allows for the inclusion of scientific insights on water governance into management practice, and vice versa.

In future studies stemming from this research, empirical research will aim to determine concrete sets of management paradigms that facilitate IWRM. In particular, the relationships of overall-system and sub-system management paradigms require additional empirical research to identify which sets of sub-system paradigms are supported or inhibited by overall-system paradigms. In addition, the applicability of the proposed methodology to design concrete pathways towards IWRM will be evaluated through further case-study research in other basins with different socio-economic and hydrological characteristics (especially transboundary basins).

## CONCLUSIONS

The analysis of management and governance systems is a complex task. Methods and tools have to deal with this complexity in order to avoid resorting to simplistic solutions or panaceas. In addition, effective science-policy-engineering dialogues need to be initiated to transfer findings between research and policy-making, and facilitate implementation on the ground.

The methodology presented in this paper complies with these requirements by building on participatory model building and the MTF. Participatory model building is a suitable and widely tested method to structure complex problems and elicit different perspectives held by stakeholders. Participatory model building supports the analysis of objective and subjective dimensions of resource issues. The MTF allows for the integration of elicited knowledge into an overall system perspective, and supports subsequent discussions of pathways towards sustainable management and governance systems. The MTF provides a formalization of complex

management and governance systems and a 'common language' that can be used in participatory processes.

Management paradigms are proposed in this paper as suitable concepts to analyze the interlinkages between resource, management and governance systems in a comprehensive way. The consideration of sub-system specific management paradigms acknowledges that the effectiveness of paradigms depends on the respective application area or sub-system. However, an overall-system paradigm usually emerges that reflects the common mindset for the overall system. Such an overall-system paradigm can support a variety of management paradigms at the sub-system level (e.g., an adaptive and integrated management paradigm), or hamper variety by being incompatible with other paradigms (e.g., a predict and control paradigm).

A case study in the Tisza River Basin, Hungary, demonstrated the application of the proposed methodology. Further empirical research will be conducted in the future to evaluate the applicability and effectiveness of the proposed methodology in different water management contexts. In addition, concrete sets of management paradigms will be examined that support integrated and adaptive water resources management.

## REFERENCES

- Brugnach, M., Dewulf, A., Pahl-Wostl, C. & Taillieu, T. 2008 Toward a relational concept of uncertainty: About knowing too little, knowing too differently, and accepting not to know. *Ecology and Society* **13** (2), 30.
- Cashman 2009 *Alternative manifestations of actor responses to urban flooding: Case studies from Bradford and Glasgow*. *Water Science and Technology* **60** (1), 77–85.
- Costanza, R. & Ruth, M. 1998 *Using dynamic modeling to scope environmental problems and build consensus*. *Environmental Management* **22** (2), 183–195.
- Folke, C. 2006 *Resilience: The emergence of a perspective for social–ecological systems analyses*. *Global Environmental Change* **16**, 253–267.
- Galaz, V. 2007 *Water governance, resilience and global environmental change – a reassessment of integrated water resources management (IWRM)*. *Water Science and Technology* **56** (4), 1–9.
- Grabs, W., Tyagi, A. C. & Hyodo, M. 2006 *Integrated flood management*. *Water Science and Technology* **56** (4), 97–103.
- Halbe, J., Pahl-Wostl, C. & Adamowski, J. (accepted) *A framework to support the initiation, planning, and institutionalization of participatory modeling processes in water resources management*. *Water Resources Research*.
- Holling, C. S. 1978 *Adaptive Environmental Assessment and Management*. Wiley, Chichester, UK.



- Jeffrey, P. & Gearey, M. 2006 Integrated water resources management: Lost on the road from ambition to realisation? *Water Science and Technology* **53** (1), 1–8.
- Jonsson, A., Andersson, L., Alkan-Olsson, J. & Arheimer, B. 2007 How participatory can participatory modeling be? Degrees of influence of stakeholder and expert perspectives in six dimensions of participatory modeling. *Water Science and Technology* **56** (1), 207–214.
- Knieper, C., Holtz, G., Kastens, B. & Pahl-Wostl, C. 2010 Analysing water governance in heterogeneous case studies – experiences with a database approach. *Environmental Science and Policy* **13** (7), 592–603.
- Medema, W., McIntosh, B. S. & Jeffrey, P. J. 2008 From premise to practice: A critical assessment of integrated water resources management and adaptive management approaches in the water sector. *Ecology and Society* **13** (2), 29.
- Metcalfe, S. S., Wheeler, E., BenDor, T., Lubinski, K. S. & Hannon, B. M. 2010 Sharing the floodplain: Mediated modeling for environmental management. *Environmental Modelling and Software* **25** (11), 1282–1290.
- Ostrom, E. 2005 *Understanding Institutional Diversity*. Princeton University Press, New Jersey, USA.
- Pahl-Wostl, C. 2007 Transitions towards adaptive management of water facing climate and global change. *Water Resources Management* **21**, 49–62.
- Pahl-Wostl, C. 2009 A conceptual framework for analysing adaptive capacity and multi-level learning processes in resource governance regimes. *Global Environmental Change* **19**, 354–365.
- Pahl-Wostl, C., Craps, M., Dewulf, A., Mostert, E., Tabara, D. & Taillieu, T. 2007 Social learning and water resources management. *Ecology and Society* **12** (2), 5.
- Pahl-Wostl, C., Holtz, G., Kastens, B. & Knieper, C. 2010 Analysing complex water governance regimes: The management and transition framework. *Environmental Science and Policy* **13** (7), 571–581.
- Pahl-Wostl, C., Jeffrey, P., Isendahl, N. & Brugnach, M. 2011 Maturing the new water management paradigm: Progressing from aspiration to practice. *Water Resources Management* **25** (3), 837–856.
- Renger, M., Kolfshoten, G. L. & de Vreede, G.-J. 2008 Challenges in collaborative modeling: A literature review. *International Journal of Simulation, and Process Modelling* **4**, 248–263.
- Schlüter, M., Hirsch, D. & Pahl-Wostl, C. 2010 Coping with change: Responses of the Uzbek water management regime to socio-economic transition and global change. *Environmental Science and Policy* **13** (7), 620–636.
- Sendzimir, J., Flachner, Z., Pahl-Wostl, C. & Knieper, C. 2010 Stalled regime transition in the upper Tisza river basin: The dynamics of linked action situations. *Environmental Science and Policy* **13** (7), 604–619.
- Sendzimir, J., Magnuszewski, P., Flachner, Z., Balogh, P., Molnar, G., Sarvari, A. & Nagy, Z. 2007 Assessing the resilience of a river management regime: Informal learning in a shadow network in the Tisza River Basin. *Ecology and Society* **13** (1), 11.
- Tidwell, V. C., Passell, H. D., Conrad, S. H. & Thomas, R. P. 2004 System dynamics modeling for community-based water planning: Application to the Middle Rio Grande. *Aquatic Science* **66** (4), 357–372.
- van den Belt, M. 2004 *Mediated Modeling – A System Dynamics Approach to Environmental Consensus Building*. Island Press, Washington.
- Voinov, A. & Bousquet, F. 2010 Modelling with stakeholders. *Environmental Modelling and Software* **25**, 1268–1281.
- Walker, W. E., Harremoës, P., Rotmans, J., van der Sluis, J. P., van Asselt, M. B. A., Janssen, P. & Krayen von Kraus, M. P. 2003 Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment* **4** (1), 5–17.

First received 28 September 2012; accepted in revised form 6 February 2013

Copyright of Water Science & Technology is the property of IWA Publishing and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

## Accepted Manuscript

Spatial and temporal trends of mean and extreme rainfall and temperature for the 33 urban centres of the arid and semi-arid state of Rajasthan, India

Santosh M. Pingale, Deepak Khare, Mahesh K. Jat, Jan Adamowski

PII: S0169-8095(13)00311-6  
DOI: doi: [10.1016/j.atmosres.2013.10.024](https://doi.org/10.1016/j.atmosres.2013.10.024)  
Reference: ATMOS 3009

To appear in: *Atmospheric Research*

Received date: 7 May 2013  
Revised date: 21 September 2013  
Accepted date: 29 October 2013



Please cite this article as: Pingale, Santosh M., Khare, Deepak, Jat, Mahesh K., Adamowski, Jan, Spatial and temporal trends of mean and extreme rainfall and temperature for the 33 urban centres of the arid and semi-arid state of Rajasthan, India, *Atmospheric Research* (2013), doi: [10.1016/j.atmosres.2013.10.024](https://doi.org/10.1016/j.atmosres.2013.10.024)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Spatial and temporal trends of mean and extreme rainfall and temperature for the 33 urban centres of the arid and semi-arid state of Rajasthan, India**

Santosh M. Pingale <sup>a\*</sup>, Deepak Khare <sup>a</sup>, Mahesh K. Jat <sup>b</sup> and Jan Adamowski <sup>c</sup>

<sup>a</sup> Department of Water Resource Development and Management,  
Indian Institute of Technology Roorkee, Roorkee 247 667 (UA), India

<sup>b</sup> Department of Civil Engineering,  
Malaviya National Institute of Technology, Jaipur, Rajasthan, India

<sup>c</sup> Department of Bioresource Engineering,  
McGill University, Quebec, Canada

Email: pingalesm@gmail.com

\*corresponding author

**ABSTRACT**

Trend analysis of the mean (monsoon, non-monsoon season and annual) and extreme annual daily rainfall and temperature at the spatial and temporal scale were carried out for all the 33 urban centres of the arid and semi-arid state of Rajasthan, India. Statistical trend analysis techniques, namely the Mann-Kendall test and Sen's slope estimator, were used to examine trends (1971-2005) at the 10% level of significance. Both positive and negative trends were observed in mean and extreme events of rainfall and temperature in the urban centres of Rajasthan State. The magnitude of the significant trend of monsoon rainfall varied from (-) 6.00 mm/hydrologic year at Nagaur to (-) 8.56 mm/hydrologic year at Tonk. However, the magnitude of the significant negative trends of non-monsoon rainfall varied from (-) 0.66 mm/hydrologic year at Dungarpur to (-) 1.27 mm/hydrologic year at Chittorgarh. The magnitude of positive trends of non-monsoon rainfall varied from 0.93 mm/hydrologic year at Churu to 1.70 mm/hydrologic year at Hanumangarh. The magnitude of the significant negative trends of annual rainfall varied from (-) 6.47 mm/year at Nagaur to (-) 10.0 mm/year at Tonk. The minimum, average and maximum temperature showed significant increasing warming trends on an annual and seasonal scale in most of the urban centres in Rajasthan State. The magnitude of statistically significant annual extreme daily rainfall varied from 2.00 mm at Jhalawar to (-) 1.64 mm at Tonk, while the magnitude of statistically significant extreme annual daily minimum and



maximum temperature varied from  $0.03^{\circ}\text{C}$  at Ganganagar to  $0.05^{\circ}\text{C}$  at Jhalawar, respectively. The spatial variations of the trends in mean (monsoon, non-monsoon season and annual) and extreme annual daily rainfall and temperature were also determined using the Inverse-Distance-Weighted (IDW) interpolation technique. IDW results are helpful to identify trends and variability in mean and extreme rainfall and temperature in space and time for the study locations where the data is not available and the quality of data is not good. These spatial maps of temperature and rainfall can help local stakeholders and water managers to understand the risks and vulnerabilities related to climate change in terms of mean and extreme events in the region.

*Key words:* Extreme events; Interpolation; Mann-Kendall test; temperature; rainfall; India

## 1. Introduction

Increasing urbanization, populations and economies are causing adverse impacts on the natural environment. With increases in urbanization, problems related to climate change are becoming worse, since additional pervious areas are becoming impervious, leading to decreases in vegetation cover and increases in temperature. Rapid economic development and population growth in various parts of the world such as India have caused concerns regarding the quantity and quality of natural resources, and in particular water resources. As such, information on the magnitude, duration and frequency of extreme events (such as floods and droughts), along with trends in various hydro-meteorological variables, is required for the planning of optimum adaptation strategies for the sustainable use of natural resources such as water (e.g. Kampata et al., 2008; Some'e et al., 2012; Huang et al., 2013). Mall et al. (2006) have studied the potential for sustainable development of surface water and groundwater resources within the constraints imposed by climate change. The authors suggested that little work has been done on hydrological impacts of climate change for Indian regions/basins. Therefore, assessment of climate change is required in India which considers the causative factors of climate change (i.e. anthropogenic factors, global warming in terms of variations in meteorological parameters) at spatial and temporal scales. Further, it is essential in hydrological modeling to assess the adverse impacts of climate change on water resources. Such studies are even more significant for urbanized watersheds/areas to better understand the relationship

between climate change and hydrological processes, which will facilitate the sustainable utilization of water resources.

There are different methods that can be used in the assessment of climate change. Various parametric and non-parametric statistical tests have been used in the recent past by many researchers to assess trends in hydro-meteorological time series in India (e.g. Gadgil and Dhorde, 2005; Singh et al., 2008a, 2008b; Basistha et al., 2009; Kumar et al., 2010; Pal and Al-Tabbaa, 2010; Dekha et al., 2012; Duhan et al., 2013; Jeganathan and Andimuthu, 2013; Jain et al., 2013), and in other regions of the world (e.g. Zhang et al., 2000; Zhang et al., 2001; Roy and Balling, 2004; Cheung et al., 2008; Motiee and McBean, 2009; Sahoo and Smith, 2009; Zin et al., 2010; Tabari et al., 2011; Nalley et al., 2012; Nemeč et al., 2012; Saboohi et al., 2012; Jiang et al. 2013). Detection of past trends, changes, and variability in the time series of hydro-climatic variables is very important for understanding the potential impact of future changes in the region (Sahoo and Smith, 2009). Statistical analysis could be extended to analyze climatic parameters and their relationships with water resources, land use/cover changes, urbanization etc.

It has been seen that the average earth's temperature has increased by  $0.6^{\circ}\text{C}$  in the later part of the 20<sup>th</sup> century; there is also a dramatic shift in temperature change from a minimum of  $1.4^{\circ}\text{C}$  to a maximum of  $5.4^{\circ}\text{C}$  as reported by the projections made by various climate prediction models (IPCC, 2001). The IPCC has demonstrated that climate change assessments at sub-national (i.e. state or province) and local scales are needed. The IPCC has also found several gaps in knowledge that exist in terms of observations and research needs related to climate change and water (IPCC, 2007). These includes such as GCM scale resolutions and downscaling techniques to locale scales to assess the climate change and to quantify the contribution of anthropogenic activity to climate change etc.

It has been found that seasonal and annual air temperature has been increasing at a rate of  $0.57^{\circ}\text{C}$  per 100 years (1881 to 1997) in India (Pant and Kumar et al., 1997). The trends of certain hydro-climatic variables have been studied in Indian urban centres such as Hyderabad, Patna, Ahmedabad, Surat, Bangalore, Mumbai, Nagpur, Pune, Jaipur, Chennai, New Delhi, Kanpur, Lucknow and Kolkata, whose populations are more than one million (De and Rao, 2004). Significant increasing trends were found in annual and monsoon rainfall over Chennai, New Delhi, Kolkata and Mumbai. Other climatic parameters were not considered to assess climate change impacts on timing, distribution, pattern, frequency, variability of precipitation and their inter-relationships in urban cities.

Arora et al. (2005) explored trends of annual average and seasonal temperature using the MK test at the country and regional scales in India. It was observed that annual mean temperature, mean maximum temperature and mean minimum temperature have increased at the rate of 0.42, 0.92 and 0.09°C (100 year)<sup>-1</sup>, respectively. On a regional basis, stations in southern and western India show a rising trend of 1.06 and 0.36°C (100 year)<sup>-1</sup>, respectively, while stations in the north Indian plains shows falling trend of 0.38°C (100 year)<sup>-1</sup>. The seasonal mean temperature has increased by 0.94°C (100 year)<sup>-1</sup> for the post-monsoon season and by 1.1°C (100 year)<sup>-1</sup> for the winter season. This study deals only with trend detection of temperature; in addition, trends in annual extreme daily events of rainfall and temperature has not been studied.

Basistha et al. (2007) assessed the spatial trends of rainfall over Indian subdivisions from 1872-2005. Their results show decreasing trends of rainfall over North India excluding Punjab, Haryana, West Rajasthan, and Saurashtra, and increased trends in south India excluding Kerala and Madhya Maharashtra. The study also concluded that the arid portion (i.e. western part) of India has not been investigated in detail. Further, they highlighted that more research is required to assess the spatial patterns of trends of other climatic variables (average, minimum and maximum temperature, relative humidity, wind speed, evapotranspiration (ET), number of rainy days etc) and their inter-relationships including trends of annual and seasonal climatic parameters to assess the impacts of climate change. Gowda et al. (2008) studied the region of Devangere district over a period of 32 years using statistical analysis. Climatic parameters (i.e. rainfall, relative humidity, maximum temperature, minimum temperature, sunshine hour and wind speed) were analyzed to assess climate change. The statistical analysis showed that such a small data set may not represent the correct picture of climate change and requires long term data. A significant change in climatic variables was found in and around the Devangere region. Ghosh et al. (2009) observed the varied trend in Indian summer monsoon rainfall, which has not only been affected by global warming, but may also be affected by local changes due to rapid urbanization, industrialization and deforestation. Patra et al. (2012) detected rainfall trends in the twentieth century in India using parametric and non-parametric statistical trend analysis tests. The temporal variation in monthly, seasonal and annual rainfall was studied for the Orissa state using data from 1871 to 2006. The analysis revealed a long term, insignificant, declining trend in annual as well as monsoon rainfall, and an increasing trend in the post-monsoon season. Rainfall during winter and summer

seasons showed an increasing trend. The study only assessed rainfall for Orissa state; trends in mean and extreme annual daily events were not explored. Duan and Pandey (2013) investigated the spatial and temporal variability of precipitation in 45 districts of Madhya Pradesh (MP), India on an annual and seasonal basis. The MK test and Sen's slope estimator test were used for trend analysis, and an increase and decrease in the precipitation trend was found on annual and seasonal basis, respectively, in the districts of MP. In this study, little attention was given to urban trends of mean and extreme rainfall and temperature. A literature review shows that in India climate related studies are restricted to statistical analysis of some of the meteorological parameters only. The effects of climate change on water resources were not considered and seasonal changes are not clearly indicated with mean and extreme events, especially in the urban centres.

As described above, the analysis of extreme trend and variability in annual daily rainfall and temperature has been carried out in recent years by a number of researchers in various parts of the world, including India. However, there is still a lack of information on trends and variability of extreme annual daily events, especially for the semi-arid and arid region of Rajasthan, India. For the region of Rajasthan, there is lack of studies of climate change assessments that consider various region specific aspects, and their inter-relation with the climatic parameters at the spatial and temporal scales. Climate change in terms of inter-seasonal and inter-annual variation in mean and extreme annual daily rainfall and temperature, for urban centres of arid and semi-arid regions in India, have not been reported in much detail in the literature. In light of this, the present study was aimed at studying climate change in terms of mean (monsoon, non-monsoon and annual) and extreme annual daily events in the 33 main urban centres of Rajasthan in India via trend analysis of two important climatic parameters (i.e. rainfall and temperature (minimum, average and maximum)) at the spatial and temporal scale. Statistical trend analysis techniques, namely the Mann-Kendall (MK) test (Mann, 1945; Kendall, 1975) and Sen's slope estimator (Sen, 1968), were used in the present study. These methods have several advantages over parametric methods (Duan and Pandey, 2013).

The paper is organized in five sections. Section two describes the study area and data used. Section three describes the methodology adopted in the study. Section four and five describes the findings of the study on climate change assessment in terms of trends in mean (monsoon, non-monsoon season and annual) and extreme annual daily rainfall and temperature in the 33 urban centres of Rajasthan state in India. Further, the spatial



variations in mean (seasonal and annual) and extreme annual daily rainfall and temperature with their trends are presented using the Inverse-Distance-Weighted (IDW) interpolation method.

## 2. Study area and data used

Rajasthan is the largest state in India, with an area of 342,000 km<sup>2</sup> of which about 60% falls under the arid category. Administratively, the state is divided into 33 districts (Fig. 1) and its climate varies from arid to semi-arid. Geographically, Rajasthan State extends from 23°3.50' to 30°14'N latitude and 69°27' to 78°19'E longitude. The study region has unique climatological characteristics (low and erratic rainfall, extremes of diurnal and annual temperatures, low humidity and high wind velocity). It also has unique physiographic characteristics. For example, the Arawali hill range is unique because it lies along the direction of the south-west monsoon, which brings rains in the region. There is also the Great Indian Desert (called the Thar desert) in the north-western part of Rajasthan, which is India's largest desert, covering almost 70% of the State; in this region there is less vegetation and the region has a significant influence on rainfall and temperature distribution across the State). The climate in the state varies from arid to semi-arid. This makes Rajasthan state different to other regions in neighboring states. The state is bordered by Gujarat, Madhya Pradesh, Uttar Pradesh, Haryana and Punjab, and is also bordered with Pakistan. The climate in Rajasthan State is characterized by low and erratic rainfall, extremes of diurnal and annual temperature, low humidity and high wind velocity. The annual average rainfall is highly variable, and it is most erratic in the western region with frequent dry spells, punctuated occasionally by heavy downpours in some years associated with passing low pressure systems over the region. The Western and Southern districts of Rajasthan State frequently experience severe droughts. Coping with low rainfall, with a high coefficient of variation across time and space, is the major challenge (Rathor, 2005).

Gridded data sets of rainfall and temperature have been used in many hydrological and climatological studies worldwide, including Australia, for hydro-climatic forecasting, climate attribution studies and climate model performance assessments (e.g., Tozer et al., 2012; Dash et al., 2013). Therefore, in the present study 0.5<sup>0</sup>×0.5<sup>0</sup> and 1<sup>0</sup>×1<sup>0</sup> gridded data sets of daily rainfall and temperature from the period 1971 to 2005 were used for the spatial and temporal variability and trend analyses in the present study. These data were

obtained from the Indian Meteorological Department (IMD), Pune. From all of India, 395 stations were used for developing the high resolution daily gridded data sets ( $1^{\circ}\times 1^{\circ}$ ) of daily temperature (minimum, maximum and mean) for the Indian region by the Indian Meteorological Department (IMD)-Pune. The modified Shepard's angular distance weighting algorithm was used for interpolating the stations of temperature data into  $1^{\circ}\times 1^{\circ}$  grids. These data sets were evaluated using the cross validation technique to estimate the interpolation error (root mean square error), which was found to be less than  $0.5^{\circ}\text{C}$ . These data sets were also compared with another (Asian Precipitation-Highly Resolved Observational Data Integration Towards the Evaluation of Water Resources (APHRODITE)) high resolution monthly data set, and a correlation of more than 0.8 was found in most parts of the country. Similarly, high resolution daily gridded data sets of rainfall ( $0.5^{\circ}\times 0.5^{\circ}$ ) were also developed (using Shepard's method), and validated using the network of 6076 rain gauge stations for the Indian region by IMD, Pune. Additional detail on the methodology is given in IMD reports (Rajeevan and Bhate, 2008; Srivastava et al., 2008).

These data sets have been found to be useful for various regional applications (such as studies of extreme temperatures, validation of numerical and climate model simulations and many environmental applications including trend analysis of mean and extreme events of rainfall and temperature) (Rajeevan and Bhate, 2008 and Srivastava et al., 2008), and they can be used for water resources planning purposes. The data was subjected to thorough quality checks (location of stations, typing error, missing data etc.) before the data analysis was performed. Hamlet et al. (2005) have reported that existing gridded data sets are not appropriate for the long term trend estimation of precipitation and temperature fields for the continental United States. The gridding procedures introduce artificial trends due to the incorporation of stations with different record lengths and locations. However, the gridded set developed by IMD Pune, addressed these issues before and after developing the gridded data sets from the observed station data before releasing the data to users. In addition, these data sets were validated with other gridded data sets developed in the region, and their application was demonstrated successfully through case studies and by other researchers (e.g., Dash et al, 2013).

Basic data quality checks (such as rejecting values, greater than exceeding known extreme values, minimum temperature, greater than maximum temperature, same temperature values for many consecutive days, unusual high values, homogeneity, location

of the station etc.) were performed for each station, and then stations were selected based on data quality or the development of the gridded data sets of temperature and rainfall. Further, homogeneity, persistence and periodicity in the data was also analysed by autocorrelation. Normalizing the time series before the analysis was performed at each urban center. The district headquarters of Rajasthan State, which has the largest population, greatest developmental activities, and administrative centres of the districts, are considered as the 'urban centres'. The geographical extent of the urban centres is defined by their municipal boundaries defined by the development authorities. The sizes and populations of the urban centres are included in Table 1. The climate of individual urban centres has been extracted from the gridded surface of temperature and rainfall corresponding to the municipal limits of the urban centres. Therefore, it preserves the meteorological characteristics of individual urban centres. The IDW assumes that the values of the variables at a point to be predicted are similar to the values of nearby observation points. The majority of the stations are located in the urban areas. Therefore, this has a major influence of urban climate. The grid covering the urban center of the district headquarters of Rajasthan State are considered to represent the climate of the urban centre. First, the grids are identified that represent the urban centre from the developed data sets, and then daily data (1971-2005) was extracted. The daily rainfall and temperature (minimum, average and maximum) data were extracted for the 33 urban centres of district headquarters in Rajasthan State (Table 1). These urban centres were selected taking into account the length and availability of records, so that most of the urban centres in the arid and semi-arid region of Rajasthan State were covered by the corresponding data to assess trends in mean (seasonal and annual) and extreme annual daily rainfall and temperature. The grids covering the urban centres were chosen for extracting the daily rainfall and temperature data, and it was assumed that the grids covering the urban centre represents the rainfall and temperature of that urban centre in Rajasthan State over a period of time (Fig. 1 and Table 1). Further, the rainfall and temperature (minimum, average and maximum) data were arranged according to monsoon, non-monsoon season and annual scale from 1971-2005 in each urban centre. The monsoon season represents the months of June to September, and the non-monsoon season represents the months of January to May and October to December.

### 3 Methodology

### 3.1. Normalisation and autocorrelation analysis of time series

Normalized rainfall and temperature time series were used to test the outliers present in the time series data using following relationship (Rai et al., 2010):

$$X_t = (x_t - \bar{x}) / \sigma \quad (1)$$

where  $X_t$  is the normalized anomaly of the series,  $x_t$  is the observed time series, and  $\bar{x}$  and  $\sigma$  are the long-term mean and standard deviation of annual/seasonal time series, respectively.

The autocorrelation test was performed to check the randomness and periodicity in the time series (Modarres and Silva, 2007). If lag-1 serial coefficients are not statistically significant then the MK test can be applied to the original time series (Luo et al., 2008; Karpouzou et al., 2010). The serial correlation coefficients of the normalized climatic series were computed for lags  $L = 0$  to  $k$ , where  $k$  is the maximum lag (i.e.  $k = n/3$ ), and  $n$  is the length of the series. The autocorrelation coefficient  $r_k$  of a discrete time series for lag- $k$  was estimated as:

$$r_k = \frac{\sum_{k=1}^{n-k} (X_t - \bar{X}_t)(X_{t+k} - \bar{X}_{t+k})}{\left[ \sum_{k=1}^{n-k} (X_t - \bar{X}_t)^2 (X_{t+k} - \bar{X}_{t+k})^2 \right]^{0.5}} \quad (2)$$

where  $r_k$  is the lag- $k$  serial correlation coefficient. The hypothesis of serial independence was then tested by lag-1 autocorrelation coefficient as  $H_0 : r_1 = 0$  against  $H_1 : |r_1| \geq 0$  using the test of significance of serial correlation (Yevjevich, 1971; Rai et al., 2010):

$$(r_k)_{t_g} = \frac{-1 \pm t_g (n - k)^{0.5}}{n - k} \quad (3)$$

where  $(r_k)_{t_g}$  is the normally distributed value of  $r_k$ , and  $t_g$  is the normally distributed statistic at 'g' level of significance. The values of  $t_g$  are 1.645, 1.965 and 2.326 at the



significance level of 0.10, 0.05 and 0.01, respectively. The null hypothesis of serial independence was rejected at the significance level  $\alpha$  (here 0.05), if  $|r_k| \geq (r_k)_{t_g}$ . For the non-normal time series data, the MK test is an appropriate choice for trend analysis (Yue and Pilon, 2004). Therefore, the MK test was used in the present study, where autocorrelation was found to be non-significant at the 5% level of significance.

### 3.2 Mann-Kendall test

The Mann-Kendall (MK) test is a non-parametric test that can be used for detecting trends in a time series (Mann, 1945) where autocorrelation is non-significant. The non-linear trend, as well as the turning point can be derived from Kendall test statistics (Kendall, 1975). This method searches for a trend in a time series without specifying whether the trend is linear or nonlinear. It has been found to be an excellent tool for trend detection and many researchers have used this test to assess the significance of trends in hydroclimatic time series data such as water quality, stream flow, temperature and precipitation (e.g., Ludwig et al., 2004; Zhang et al., 2004; McBean and Motiee, 2008; Basistha et al., 2009; Rai et al., 2010; Patra et al., 2012). The MK test can be applied to a time series  $x_i$  ranked from  $i = 1, 2, \dots, n-1$  and  $x_j$  ranked from  $j = i+1, 2, \dots, n$  such that:

$$\text{sgn}(x_j - x_i) = \begin{cases} 1 & \text{if } (x_j - x_i) > 0 \\ 0 & \text{if } (x_j - x_i) = 0 \\ -1 & \text{if } (x_j - x_i) < 0 \end{cases} \quad (4)$$

The Kendall test statistic  $S$  can be computed as:

$$S = \sum_{k=1}^{n-1} \text{sgn}(x_j - x_k) \quad (5)$$

where  $\text{sgn}(x_j - x_k)$  is the signum function. The test statistic  $S$  is assumed to be asymptotically normal, with  $E(S) = 0$  for the sample size  $n \geq 8$  and variance as follows:

$$V(S) = \frac{[n(n-1)(2n+5) - \sum_t t(t-1)(2t+5)]}{18} \quad (6)$$

where  $t_i$  denotes number of ties upto sample  $i$

The standardized MK test statistic ( $Z_{mk}$ ) can be estimated as:

$$Z_{mk} = \begin{cases} \frac{S-1}{\sqrt{V(S)}} & \text{if } S > 0 \\ 0 & \text{if } S = 0 \\ \frac{S+1}{\sqrt{V(S)}} & \text{if } S < 0 \end{cases} \quad (7)$$

The standardized MK test statistic ( $Z_{mk}$ ) follows the standard normal distribution with a mean of zero and variance of one. If  $\pm Z_{mk} \leq Z_{\alpha/2}$  (here  $\alpha = 0.1$ ), then the null hypothesis for no trend is accepted in a two sided test for trend, and the null hypothesis for the no trend is rejected if  $\pm Z_{mk} \geq Z_{\alpha/2}$ . Failing to reject  $H_0$  (i.e. null hypothesis) does not mean that there is no trend. Rather, it is a statement that the evidence available is not sufficient to conclude if there is a trend (Helsel et al., 2002). A positive value of  $Z_{mk}$  indicates an ‘upward trend’ and a negative value indicates a ‘downward trend’. The significance levels (p-values) for each trend test can be obtained from the relationship given as (Coulibaly et al., 2005):

$$p = 0.5 - \phi(|Z_{mk}|) \quad (8)$$

where  $\phi(\cdot)$  denotes the cumulative distribution function (CDF) of a standard normal variate. At a significance level of 0.1, if  $p \leq 0.1$  then the existing trend is considered to be statistically significant.

### 3.3 Sen’s estimator of slope and percentage change over a period

If a linear trend is present in a time series, then the true slope of the trend is estimated using a simple non-parametric procedure (Theil, 1950; Sen, 1968), which is given by:

$$Q_i = \text{Median} \left( \frac{x_j - x_k}{j - k} \right) \quad \forall k \leq j \quad (9)$$

where  $x_j$  and  $x_k$  are data values at times  $j$  and  $k$  ( $j > k$ ), respectively. The median of  $N$  values of  $Q_i$  is Sen's estimator of slope. If  $N$  is odd, then Sen's estimator is computed by  $Q_{\text{med}} = Q_{(N+1)} / 2$  and if  $N$  is even, then Sen's estimator is computed by  $Q_{\text{med}} = [Q_N / 2 + Q_{(N+2)} / 2] / 2$ . Finally,  $Q_{\text{med}}$  is tested by a two-sided test at  $100(1 - \alpha)\%$  confidence interval, and the true slope is obtained by a non-parametric test.

The percentage change over a period of time can be estimated from Theil and Sen's median slope and mean by assuming the linear trend (Yue and Hashino, 2003; Basistha et al., 2009) in the time series:

$$\% \text{ change} = \left( \frac{\text{Median Slope} \times \text{length of period}}{\text{mean}} \right) \quad (10)$$

### 3.4 Interpolation Technique

In the present study, the spatial variation and trends of the mean (monsoon, non-monsoon and annual) and extreme annual daily rainfall and temperature was determined using the Inverse Distance Weighted (IDW) interpolation technique. The IDW method is a simple and effective interpolation technique that is based on the assumption that the values of the variables at a point to be predicted are similar to the values of nearby observation points. This method implies that each station has a local influence, which decreases with distance by means of the use of a power parameter (Isaaks and Srivastava, 1989; Robinson and Metternicht, 2006). The general formula of IDW (Yavuz and Erdogan, 2012) is as follows:

$$\hat{Z}(S_0) = m \sum_{i=1}^N \lambda_i Z(S_i) \quad (11)$$

where

$\hat{Z}(S_0)$  is the value of prediction for location  $S_0$

$N$  is the number of measured sample points surrounding the prediction location

$\lambda_i$  are the weights assigned to each measured point

$Z(S_i)$  is the observed value at the location  $S_i$

The weighting is a function of inverse distance. The formula determining the weighting is (Yavuz and Erdogan, 2012):

$$\lambda_i = \frac{d_{io}^{-p}}{\sum_{i=1}^N d_{io}^{-p}} \quad (12)$$

The IDW method requires the choice of a power parameter and a search radius. The power parameter ( $p$ ) controls the significance of measured values on the interpolated value based upon their distance from the output point ( $d$ ) (Erdogan 2009). The choice of a relatively high power parameter ensures a high degree of local influence and gives the output surface in more detail (Yavuz and Erdogan, 2012).

#### 4 Results and Discussion

Detailed trend analysis of the mean (monsoon, non-monsoon season and annual) and extreme annual daily rainfall and temperature was carried out for the period of 1971-2005 using the Mann-Kendall test for all 33 urban centres of Rajasthan State. Similarly, Sen's slope and percentage change in rainfall and temperature was also determined. The corresponding spatial and temporal distribution of the mean (monsoon, non-monsoon season and annual) and extreme events of annual daily rainfall and temperature using the MK test statistics ( $Z_{mk}$ ) was determined by the IDW interpolation technique. The increasing, decreasing, and no trend at 10% level of significance are represented with the symbols  $\uparrow$ ,  $\downarrow$  and  $\circ$ , respectively.

##### 4.1 Trend Analysis of Mean Events

###### 4.1.1 Rainfall

The spatial distributions of trends in seasonal (monsoon and non-monsoon) and annual average rainfall for the 33 urban centres of Rajasthan State are shown in Fig. 2. The magnitude of the Sen's slope and percentage change in average rainfall at seasonal (monsoon and non-monsoon) and annual temporal scales for the 33 different urban centres of Rajasthan State are also given in Tables S1 to S3 (Supporting material), respectively.



Both positive and negative trends were observed by the MK test and Sen's Slope estimator in seasonal and annual average rainfall in the urban centres. This is in good agreement with observations in different parts of India (Pal and Al-Tabbaa, 2011).

All of the significant trends in monsoon and annual scale were found to be decreasing at the 10% level of significance. However, both positive and negative trends in non-monsoon average rainfall were observed. As a result, the expected decrease in rainfall will possibly cause a decrease in water availability in the future (Gosain et al., 2006; Tabari and Talaei, 2011b). The monsoon season rainfall pattern and trends were very similar to the annual ones, which demonstrate the major contribution of monsoon rainfall in annual rainfall in the urban centres of Rajasthan State. Similar results indicating the major contribution of seasonal precipitation in annual precipitation has also been observed in Iran (Tabari and Talaei, 2011b). Significant decreasing trends were observed in monsoon average rainfall for Jodhpur (at -6.82 mm/hydrologic yr and -88.16%), Karauli (at -7.59 mm/hydrologic yr and -39.03%), Nagaur (at -6.00 mm/hydrologic yr and -66.92%) and Tonk (at -8.56 mm/hydrologic yr and -50.63%). The significant decreasing trends of average monsoon rainfall had the highest value of decreasing slope (i.e. -8.56 mm/hydrologic year) for Tonk, and percentage change of -88.16% in Jodhpur (Table S1). Significant trends in non-monsoon average rainfall were found for Dungarpur (at -0.66 mm/hydrologic yr and -58.10%), Chittorgarh (at -1.27 mm/hydrologic yr and -80.29%), Jaipur (at 1.22 mm/hydrologic yr and 61.53%), Churu (at 0.93 mm/hydrologic yr and 49.61%), Ganganagar (at 1.39 mm/hydrologic yr and 79.46%) and Hanumangarh (at 1.70 mm/hydrologic yr and 102.85%). A significant increasing trend was observed in non-monsoon rainfall at Hanumangarh (1.70 mm/hydrologic year), and a decreasing trend was found at Chittorgarh (-1.27 mm/hydrologic year) (Table S2). The percentage change of significant trend was found to be 102.85% and -80.29% in Hanumangarh and Chittorgarh, respectively (Table S2).

Significant decreasing trends were observed in annual average rainfall for Jodhpur (at -7.51 mm/yr and -85.05%), Karauli (at -7.41 mm/yr and -34.65%), Nagaur (at -6.47 mm/yr and -61.93%) and Tonk (at -10 mm/yr and -54.54%). In the case of annual average rainfall, maximum values of decreasing slopes (-10 mm/year) and percentage change (-85.05%) were observed in Tonk and Jodhpur, respectively (Table S3). The IPCC also reported increases in precipitation at a global level, however, in the present study, precipitation was found to increase in some urban centres and decrease in other centres.

This may be due to changing land use/land cover, and several other anthropogenic activities at a local scale (Gowda et al., 2008).

#### 4.1.2 Minimum temperature

Significant increasing trends were observed in minimum temperature at seasonal and annual temporal scales for the majority of the urban centres in the Northeastern part of Rajasthan State. However, significant trends were not found in other urban centres at 10% level of significance by the MK test (Fig. 3). The magnitude of the Sen's slope of significant trends in seasonal (monsoon, non-monsoon and annual) minimum temperature for the urban centres of Rajasthan State is shown in Fig. 3. The maximum value of Sen's slope ( $0.04^{\circ}\text{C}/\text{hydrologic year}$ ) and percentage change (5.07%) for a significant increasing trend of average minimum temperature in the monsoon season were observed in Churu and Sikar, respectively. In the case of non-monsoon average minimum temperature, maximum values of increasing ( $0.05^{\circ}\text{C}/\text{hydrologic year}$ ) slope were observed in Bikaner, Jaipur and Sikar. Similarly, percentage change in non-monsoon minimum temperature varied from 3.91% at Bharatpur to 13.45% at Sikar. Also, a significant trend in annual average minimum temperature had maximum values of increasing slopes ( $0.05^{\circ}\text{C}/\text{year}$ ) at Sikar. The percentage changes in annual average minimum temperature were found to be at a minimum (2.61%) and maximum (9.52%) at Banaswara and Sikar, respectively. The positive trends of seasonal and annual minimum temperature were also observed by other researchers in India (e.g., Jain et al., 2013).

#### 4.1.3 Average temperature

Significant increasing trends in average temperature at seasonal and annual temporal scales were observed for the majority of the urban centres in Rajasthan State, accounting for an average 80% of the urban centres (Fig. 4). The significant increasing trends in monsoon average temperature were observed in Bikaner, Bundi, Churu, Dausa, Jaipur, Jaisalmer, Kota and Sikar, which is indicated by the maximum value of slope (i.e.  $0.03^{\circ}\text{C}/\text{hydrologic year}$ ). Also, percentage change in monsoon average temperature was found to be at a maximum for Jaisalmer (3.47%) and at a minimum for Alwar (2.43%). Similarly, significant trends in non-monsoon and annual average temperature showed the maximum value of  $0.03^{\circ}\text{C}$  in the urban centres. The percentage changes in significant trends in non-monsoon average temperature were observed to be at a minimum (2.43%)

and maximum (4.84%) in Banswara and Sikar, respectively. The percentage change in annual average temperature was found to be at a minimum (1.53%) and maximum (4.32%) in Dungarpur and Sikar, respectively. The detail magnitude of the Sen's slope of the significant trends in average temperature is shown in Fig. 4. This warming trend in average temperature is also supported by other studies in India (e.g., Singh et al., 2008).

#### 4.1.4 Maximum temperature

Significant increasing trends in monsoon average maximum temperature were found in Barmer, Bikaner, Dausa, Jaisalmer, Jalore and Pali (Fig. 5a). The significant increasing trends in monsoon season average maximum temperature were observed with a maximum value ( $0.03^{\circ}\text{C}/\text{hydrologic year}$ ) of Sen's slope in Jaisalmer and Bikaner. Percentage change in average maximum temperature during the monsoon period was at a minimum (2.09%) and maximum (2.90%) in Jalore and Bikaner, respectively. In the case of non-monsoon average maximum temperature, the Southwestern urban centres in Rajasthan State had significant increasing trends. No significant trend was found in the remaining urban centres during the non-monsoon period at 10% level of significance (Fig. 5b). Significant increasing trends were observed in annual average maximum temperature for urban centres that are located in the Southeast and Southwestern part of Rajasthan State (Fig. 5c). The significant increasing trends in non-monsoon season average maximum temperature were observed with a maximum value of Sen's slopes ( $0.03^{\circ}\text{C}/\text{hydrologic year}$ ) in Barmer, Jaisalmer, Jalore, Pali, Sirohi and Udaipur. Percentage changes in non-monsoon average maximum temperature were found to be at a minimum (2.08%) and maximum (3.31%) in Dungarpur and Pali, respectively. No significant trends were found in the average maximum temperature at an annual temporal scale for the urban centres located in the Northern region of the State (Fig. 5c). Similarly, significant trends in annual average maximum temperature were observed with maximum values of slopes ( $0.03^{\circ}\text{C}/\text{year}$ ) in Barmer, Jaisalmer, Jalore, Pali and Udaipur. The percentage change was found to be at a minimum (1.47%) and maximum (3.24%) in Pali and Baran, respectively. The detail magnitude of the Sen's slope for average maximum temperature for the urban centres of Rajasthan State is shown in Fig. 5. These results are in good agreement with previous studies in India (e.g., Singh et al., 2008).

#### 4.2 Trends in Extreme Events

#### 4.2.1 Extreme annual daily rainfall

The spatial distribution of extreme annual maximum daily rainfall trends and variation are shown in Fig. 6(a), while the MK test statistics ( $Z_{mk}$ ) with increasing, decreasing and no trends are shown in Fig. 6(b). Both positive (Southeastern) and negative (Northeastern) trends were observed in extreme annual daily rainfall in the urban centres of Rajasthan State. The results revealed that significant decreasing trends and variability were observed in annual extreme daily rainfall for Hanumangarh (at -0.65 mm/yr and -49.16%), Karauli (at -1.56 mm/yr and -60.31%), Nagaur (at -1.56 mm/yr and -80.33%), Sikar (at -1.33 mm/yr and -63.86%) and Tonk (at -1.64 mm/yr and -65.14%). However, a significant increasing trend was found only in Jhalawar (at 1.38 mm/yr and 45.97%). The maximum values of decreasing (-1.64 mm/year) slopes and percentage change (-80.33%) in significant trends in annual maximum daily rainfall were observed in Tonk and Nagaur, respectively (Table S4). Significant trends were not found for extreme events of annual daily rainfall in other urban centres (Fig. 6). The detail magnitudes of slope and percentage change in extreme annual maximum daily rainfall are presented in Table S4 and Fig. 6. Similarly, only significant trends in extreme annual maximum daily rainfall are shown in Fig. 7. Investigating trends in extreme events is necessary to assess the impact of global climate change over regional climatic variability on present and future water resources (Wagesho et al., 2013). The assessment of trends of extreme annual daily rainfall events is crucial for adaptation planning in arid and semi-arid regions of India such as Rajasthan.

#### 4.2.2 Extreme annual daily minimum temperature

The spatial distribution of extreme annual daily minimum temperature is shown in Fig. 8(a), while the MK test statistic ( $Z_{mk}$ ) indicating trends is shown in Fig. 8(b). Significant increasing trends were observed in extreme annual daily minimum temperature (at 0.03 to 0.07<sup>0</sup>C/yr) for the urban centres located in the Northeastern and Western region of Rajasthan State (Fig. 8). However, significant trends were not found in other urban centres of Rajasthan State (Fig. S1) (Supporting material). A significant increasing trend in annual minimum daily temperature was observed with a maximum value of slope (0.073<sup>0</sup>C/year) and percentage change (102.76%) in Sikar, while significant increasing annual daily minimum temperature was observed with a minimum value of slope (0.040<sup>0</sup>C/year) and percentage change (31.01%) in Banswara and Hanumangarh,



respectively. The details of slope magnitude and variability of extreme annual daily minimum temperature in the urban centres of Rajasthan State are presented in Table S5 and Fig. 8. Only the trends in significant extreme events of annual daily minimum temperature are shown in Fig. S1. From the results, it is clear that the magnitude of the minimum temperature trends and variability is greater than the maximum temperature extremes. This was also observed in other parts of India (Duhan et al., 2013).

#### 4.2.3 Extreme annual daily maximum temperature

The spatial distribution of extreme annual daily maximum temperature (obtained using MK test statistic ( $Z_{mk}$ )) was determined using the IDW interpolation technique for the 33 urban centres of Rajasthan State (Fig. 9). A significant increasing trend in extreme annual daily maximum temperature was observed in the South (i.e. Banaswara, Baran, Bundi, Chittorgarh, Jhalawar, Pali and Sirohi), East (i.e. Churu and Jaipur) and West (i.e. Jaisalmer) urban centres of Rajasthan State (Fig. 9). Significant increasing trend in the annual maximum daily temperature was observed with a maximum value of Sen's slope ( $0.049^{\circ}\text{C}/\text{year}$ ) and percentage change (3.86%) in Jhalawar, while the minimum value of slope ( $0.029^{\circ}\text{C}/\text{year}$ ) and percentage change (2.35%) was observed in Sirohi. The minimum and maximum annual daily maximum temperature was found to be  $44.48^{\circ}\text{C}$  and  $47.62^{\circ}\text{C}$  at Sirohi and Dholpur, respectively. The detail magnitude of the Sen's slope and percentage change of annual daily maximum temperature in urban centres of Rajasthan State are presented in Table S6. The results of significant extreme events of annual daily maximum temperature are shown in Fig. S2. The results clearly indicate the warming trends in the extreme annual maximum daily temperature in the urban centres, which shows attribution of anthropogenic influence on climate (Fig. 9a and b). These may include drastic changes in land use/land cover, industrialization and urbanisation.

In summary, the mean and extreme rainfall and temperature trends and variability at a spatial and temporal scale were evaluated for the 33 urban centres of Rajasthan State. The results of mean and extreme annual daily rainfall showed both positive and negative trends over the urban centres. An increasing trend was found in annual and seasonal mean and extreme minimum, average and maximum temperature over the urban centres. This may be due to several natural and anthropogenic activities such as land use/land cover, urbanisation and industrialization, which contribute to the changes in rainfall and temperature in terms of their distribution, intensity and occurrence in space and time. The

detailed evaluation of the magnitude and variability in rainfall, and minimum and maximum temperature in terms of mean and extremes, is very important for impact assessment and adaptation planning for floods, drought and extreme events. It is suggested that in the future, the assessment of mean and extreme events considering climate change scenarios can be carried out on seasonal and annual temporal scale at each urban centres of Rajasthan State, India. This would be very useful to facilitate the proper planning and management of water resources in a changing climate in Rajasthan State, India.

#### 4 Conclusions

In the present study, the trends in mean (monsoon, non-monsoon season and annual) and extreme annual daily rainfall and temperature were determined for the 33 urban centres of Rajasthan State. This was carried out using the non-parametric Mann-Kendall (MK) test. The Sen's slope and percentage changes in rainfall and temperature were also estimated over the study period (1971-2005). Further, the spatial variations in mean (seasonal and annual) and extreme annual daily rainfall and temperature with their trends were determined using the Inverse-Distance Weighted (IDW) interpolation method.

Both positive and negative trends were observed by the MK test and Sen's Slope estimator in seasonal and annual average rainfall in the urban centres of Rajasthan State. All of the significant trends in monsoon and annual rainfall were found to be decreasing at the 10% level of significance. However, significant trends in non-monsoon average rainfall were observed (both positive and negative trends). The significant decreasing trends of average monsoon rainfall had the highest value of decreasing slope (i.e. -8.56 mm/hydrologic year) for Tonk and percentage change of -88.16% in Jodhpur. A significant increasing trend was observed in non-monsoon rainfall at Hanumangarh (1.70 mm/hydrologic year), and a decreasing trend was found at Chittorgarh (-1.27 mm/hydrologic year). In the case of annual average rainfall, maximum values of decreasing slopes (-10 mm/year) and percentage change (-85.05%) were observed in Tonk and Jodhpur, respectively. Significant increasing trends were observed in minimum temperature at seasonal (0.04 and 0.05<sup>0</sup>C/hydrologic year) and annual (0.05<sup>0</sup>C/year) temporal scales for the majority of the urban centres in the Northeastern part of Rajasthan State. Significant increasing trends (at 0.01 to 0.05<sup>0</sup>C/yr) were observed in average temperature for most of the urban centres (about 80% of the urban centres), in particular, all the urban centres showed significant warming trends in annual average temperature.

Significant increasing trends in monsoon average maximum temperature were found in the some of the urban centres. However, the Southwestern urban centres had significant increasing trends in non-monsoon average maximum temperature. Significant increasing trends were observed in annual average maximum temperature for urban centres that are located in the Southeast and Southwestern part of Rajasthan State. The trends in maximum and minimum temperature may be influenced in different ways due to local physical geographic (e.g. topography, urbanization and air pollution etc.) and atmospheric circulation features (Turkes and Sumer, 2004; Dhorde et al., 2009; Tabari and Talae, 2011a).

In the case of extreme events for annual daily rainfall both positive (Southeastern) and negative (Northeastern) trends were observed in some of the urban centres. Other urban centres did not show any significant trends in extreme annual daily rainfall. Significant increasing trends were observed in extreme annual daily minimum temperature (at 0.03 to 0.07<sup>0</sup>C/yr) for the urban centres located in the Northeastern and Western region of Rajasthan State. However, significant trends were not found in the other urban centres of Rajasthan State. A significant increasing trend in extreme annual daily maximum temperature (at 0.03 to 0.05<sup>0</sup>C/year) was observed in the South (i.e. Banaswara, Baran, Bundi, Chittorgarh, Jhalawar, Pali and Sirohi), East (i.e. Churu and Jaipur) and West (i.e. Jaisalmer) urban centres of Rajasthan State. The results of mean (monsoon, non-monsoon and annual) and extreme annual daily rainfall and temperature clearly indicates prominent changes in the urban centres, whether due to natural variability or anthropogenic activity over the period. The trends and variability of rainfall and temperature found in this study could be one part of a cycle with a frequency shorter than 35 years. A similar observation was also made for precipitation trends in Iran (Tabari and Talae, 2011a). There are no doubt changes in natural factors like sea surface temperature, the earth's rotation and solar cycles that can cause changes in the climate. However, we do not have any control of these factors. Anthropogenic activities like urbanisation and land use change can also affect local climates significantly. The effects of these factors could be minimised by taking appropriate adaptation and mitigation measures. The assessment of mean and extreme events of rainfall and temperature are necessary to help and prepare suitable adaptation strategies in the face of uncertain changing climate and extreme weather events. The outcome of this study should prove to be useful to policy makers, hydrologists, and

water resources planners dealing with climate change in all the urban centres in Rajasthan State for the sustainable development and planning of water resources.

### **Acknowledgements**

We acknowledge the Indian Meteorological Department (IMD) in Pune for providing useful meteorological gridded datasets for this research work. The financial assistance provided by Ministry of Human Resource Development, Government of India in the form of a scholarship is duly acknowledged.

### **References**

- Arora, M., Goel, N.K., Singh, P., 2005. Evaluation of temperature trends over India. *Hydrological Sciences Journal*, 50(1), 81-93.
- Basistha, A., Arya, D.S., Goel, N.K., 2009. Analysis of historical changes in rainfall in the Indian Himalayas. *International Journal of Climatology* 29, 555–572.
- Basistha, A., Goel, N.K., Arya, D.S., Gangwar, S.K., 2007. Spatial pattern of trends in Indian sub-divisional rainfall. *Jalvigyan Sameeksha* 22, 47-57.
- Cheung, W.H., Senay, G.B., Singh, A., 2008. Trends and spatial distribution of annual and seasonal rainfall in Ethiopia. *International Journal of Climatology* 28, 1723-1734.
- Coulibaly P., Shi, X., 2005. Identification of the effect of climate change on future design standards of drainage infrastructure in Ontario, final report, McMaster University, Department of Civil Engineering, Ontario.
- Dash, S.K., Saraswat, Vaishali, Panda, S.K., Sharma, N., 2013. A Study of Changes in Rainfall and Temperature Patterns at Four Cities and corresponding Meteorological Subdivisions over Coastal Regions of India, *Global and Planetary Change*, doi: 10.1016/j.gloplacha.2013.06.004.
- De, U.S., Rao, G.S.P., 2004. Urban climate trends - The Indian scenario. *Journal of Indian Geophysical Union* 8(3), 199-203.
- Deka, R.L., Mahanta, C., Pathak, H., Nath, K.K., Das, S., 2012. Trends and fluctuations of rainfall regime in the Brahmaputra and Barak basins of Assam, India. *Theor Appl Climatol.*, DOI 10.1007/s00704-012-0820-x.
- Dhorde, A., Dhorde, A., Gadgil, A.S., 2009. Long-term temperature trends at four largest cities of India during the twentieth century. *J. Ind. Geophys. Union*. 13 (2), 85–97.



- Duhan, D., Pandey, A., 2013. Statistical analysis of long term spatial and temporal trends of precipitation during 1901–2002 at Madhya Pradesh, India. *Atmospheric Research* 122, 136–149.
- Duhan, D., Pandey, A., Gahalaut, K.P.S., Pandey, R.P., 2013. Spatial and temporal variability in maximum, minimum and mean air temperatures at Madhya Pradesh in central India. *C. R. Geoscience* 345, 3–21.
- Erdoğan S., 2009. A comparison of interpolation methods for producing digital elevation models at the field scale. *Earth Surf Process Land* 34, 366–376.
- Gadgil, A., Dhorde, A., 2005. Temperature trends in twentieth century at Pune, India. *Atmos. Environ.*, 39, 6550–6556.
- Ghosh, S., Luniya, V., Gupta, A., 2009. Trend analysis of Indian summer monsoon rainfall at different spatial scales. *Atmospheric Science Letters* 10, 285-290.
- Gosain, A.K., Rao, S., Basuray, D., 2006. Climate change impact assessment on hydrology of Indian river basins. *Current Science* 90, 346–353.
- Gowda, K.K., Majuantha, K., Manjunath, B.M., Putty, Y.R., 2008. Study of climate changes at Davangere region by using climatological data. *Water and Energy International* 65(3), 66-77.
- Hamlet, A.F., Mote, P.W., Clark, M.P., Lettenmaier, D.P., 2005. Effects of Temperature and Precipitation Variability on Snowpack Trends in the Western United States. *Journal of Climate* 18, 4545-4561.
- Helsel, D.R., Hirsch, R.M., 2002. Statistical methods in water resources techniques of water resources investigations. Book 4, chapter A3. U.S. Geological Survey. 522 pages.
- Huang, J., Sun, S., Zhang, J., 2013. Detection of trends in precipitation during 1960–2008 in Jiangxi province, southeast China, *Theor Appl Climatol*, DOI 10.1007/s00704-013-0831-2.
- IPCC, 2001. Third Assessment Report, Vol. 1 “The Scientific Basis” Cambridge Univ. Press.
- IPCC, 2007. AR4 Working group I report “The physical science basis”.
- Isaaks, E.H., Srivastava, R.M., 1989. An introduction to applied geostatistics. Oxford University Press, New York.
- Jain, S.K., Kumar, V., Saharia, M., 2013. Analysis of rainfall and temperature trends in northeast India. *Int. J. Climatol* 33, 968–978.

- Jeganathan, A., Andimuthu, R., 2013. Temperature trends of Chennai City, India. *Theor Appl Climatol* 111, 417–425.
- Jiang, F., Hu, R., Wang, S., Zhang, Y., Tong, L., 2013. Trends of precipitation extremes during 1960–2008 in Xinjiang, the Northwest China. *Theor Appl Climatol* 111, 133–148.
- Kampata, J.M., Parida, B.P., Moalafhi, D.B., 2008. Trend analysis of rainfall in the headstreams of the Zambezi River Basin in Zambia. *Physics and Chemistry of the Earth* 33, 621–625.
- Karpouzou, D.K., Kavalieratou, S., Babajimopoulos, C., 2010. Trend analysis of precipitation data in Pieria region (Greece). *European Water* 30, 31-40.
- Kendall, M.G., 1975. *Rank Correlation Methods*. 4<sup>th</sup> ed., Charles Griffin, London.
- Kumar, V., Jain, S.K., Singh, Y., 2010. Analysis of long-term rainfall trends in India. *Hydrological Sciences Journal* 55(4), 484-496.
- Ludwig, W., Serrat, P., Cesmat, L., Esteves, J.G., 2004. Evaluating the impact of the recent temperature increase on the hydrology of the T<sup>et</sup> River (Southern France). *Journal of Hydrology* 289, 204–221.
- Luo, Y., Lio, S., Sheng, L., Fu, S., Liu, J., Wang, G., Zhou, G., 2008. Trends of precipitation in Beijiang River Basin, Guangdong Province, China. *Hydrological Processes* 22, 2377–2386.
- Mall, R.K., Gupta, A., Singh, R., Singh, R.S., Rathore, L.S., 2006. Water resources and climate change: An Indian perspective. *Current Science* 90 (12), 1610-1626.
- Mann, H.B., 1945. Non-parametric test against trend. *Econometrica* 13, 245–259.
- McBean, E., Motiee, H., 2008. Assessment of impacts of climate change on water resources: a long term analysis of the Great Lakes of North America. *Hydrology and Earth System Sciences* 12, 239-255.
- Modarres, R., da Silva, R.V.P., 2007. Rainfall trends in arid and semi-arid regions of Iran. *Journal of Arid Environment* 70, 344–355.
- Motiee, H., McBean, E., 2009. An Assessment of Long-Term Trends in Hydrologic Components and Implications of Water Levels in Lake Superior, *Hydrology Research Journal*, 40(6), 564-579.
- Nalley, D., Adamowski, J., Khalil, B., 2012. Using discrete wavelet transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954–2008). *Journal of Hydrology* 475, 204–228.

- Nemec, J.C, Gruber, B., Chimani, Auer, I., 2012., Trends in extreme temperature indices in Austria based on a new homogenized dataset. *Int. J. Climatol.*, 33, 1538–1550.
- Pal, I. Al-Tabbaa, A., 2011. Assessing seasonal precipitation trends in India using parametric and non-parametric statistical techniques. *Theor Appl Climatol* 103, 1–11.
- Pal, I., Al-Tabbaa, A., 2010. Long-term changes and variability of monthly extreme temperatures in India. *Theor Appl Climatol* 100, 45–56.
- Pant, G.B., Kumar, K.R., 1997. *Climates of South Asia*, John Wiley, UK, pp-320.
- Patra, J.P., Mishra, A., Singh, R., Raghuvanshi, N.S., 2012. Detecting rainfall trends in twentieth century (1871-2006) over Orissa State, India. *Climatic Change* 111(3), 801-817.
- Rai, R.K., Upadhyay, A., Ojha, C.S.P., 2010. Temporal variability of climatic parameters of Yamuna river basin: spatial analysis of persistence, trend and periodicity. *The Open Hydrology Journal* 4, 184-210.
- Rajeevan, M., Bhate J., 2008. A High Resolution Daily Gridded Rainfall Data Set (1971-2005) for Mesoscale Meteorological Studies, National Climate Centre, India Meteorological Department, Pune, Research Report No: 8/2008.
- Rathor, M.S., 2005. State level analysis of drought policies and impacts in Rajasthan, India. Colombo, Sri Lanka: IWMI. 40p. Working paper 93: Drought Series Paper 6.
- Robinson, T.P., Metternicht, G., 2006. Testing the performance of spatial interpolation techniques for mapping soil properties. *Comput. Electron. Agr.* 50, 97–108.
- Roy, S.S., Balling, R.C., Jr., (2004). Trend in extreme daily precipitation indices in India. *International Journal of Climatology* 24, 457–466.
- Saboohi, R., Soltani, S., Khodagholi, M., 2012. Trend analysis of temperature parameters in Iran. *Theor Appl Climatol* 109, 529–547.
- Sahoo, D., Smith, P.K., 2009. Hydroclimatic trend detection in a rapidly urbanizing semi-arid and coastal river basin. *Journal of Hydrology* 367, 217-227.
- Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association* 39, 1379–1389.
- Singh, P., Kumar, V., Thomas T., Arora, M., 2008a. Basin-wide assessment of temperature trends in northwest and central India. *Hydrological Sciences Journal* 53(2), 421-433.
- Singh, P., Kumar, V., Thomas T., Arora, M., 2008b. Changes in rainfall and relative humidity in river basins in northwest and central India. *Hydrol. Process.* 22, 2982–2992.

- Some'e, B.S., Ezani, A., Tabari, H., 2012. Spatiotemporal trends and change point of precipitation in Iran. *Atmospheric Research* 113, 1–12.
- Srivastava, A.K., Rajeevan, M., Kshirsagar, S.R., 2008. Development of a High Resolution Daily Gridded Temperature Data Set (1969-2005) for the Indian Region, National Climate Centre, India Meteorological Department, Pune, Research Report No: 8/2008.
- Tabari, H., Somee, B.S., Zadeh, R.M., 2011. Testing for long-term trends in climatic variables in Iran. *Atmospheric Research* 100, 132–140.
- Tabari, H., Talaei, P.H., 2011a. Analysis of trends in temperature data in arid and semi-arid regions of Iran. *Global and Planetary Change* 79, 1–10.
- Tabari, H., Talaei, P.H., 2011b. Temporal variability of precipitation over Iran: 1966–2005. *Journal of hydrology* 396, 313–320.
- Theil, H., 1950. A rank-invariant method of linear and polynomial regression analysis. I, II, III. *Nedal. Akad. Wetensch. Proc.* 53, 386–392.
- Tozer, C.R., Kiem, A.S., Verdon-Kidd, D.C., 2012. On the uncertainties associated with using gridded rainfall data as a proxy for observed. *Hydrology and Earth System Sciences*. 16, 1481–1499.
- Turkes, M., Sumer, U.M., 2004. Spatial and temporal patterns of trends and variability in diurnal temperature ranges of Turkey. *Theor. Appl. Climatol.* 77, 195–227.
- Wang, H., Chen, Y., Xun, S., Lai, D., Fan, Y., Li, Z., 2013. Changes in daily climate extremes in the arid area of northwestern China. *Theor Appl Climatol* 112, 15–28.
- Yavuz, H., Erdogan, S., 2012. Spatial analysis of monthly and annual precipitation trends in Turkey. *Water Resour Manag* 26(3), 609–621.
- Yevjevich, V., 1971. *Stochastic Processes in Hydrology*. Water Resources Publications, Fort Collins: CO.
- Yue, S., Hashino, M., 2003. Long term trends of annual and monthly precipitation in Japan. *Journal of the American Water Resources Association* 39(3), 587–596.
- Yue, S., Pilon, P., 2004. A comparison of the power of the t test, Mann-Kendall and bootstrap tests for trend detection. *Hydrological Sciences Journal* 49(1), 21–37.
- Zhang, X., Hogg, W.D., Mekis, F., 2001. Spatial and temporal characteristics of heavy precipitation events over Canada. *Journal of Climate* 14, 1923–1936.
- Zhang, X., Vincent, L.A., Hogg, W.D., Niitsoo A., 2000. Temperature and precipitation trends in Canada during the 20th century. *Atmosphere-Ocean* 38(3), 395–429.



Zhang, X.B., Zwiers, F.W., Li, G.L., 2004. Monte Carlo experiments on the direction of trends in extreme values. *Journal of Climate* 17, 1945–1952.

Zin, W.Z.W., Jamaludin, S., Deni S.M., Jemain, A.A., 2010. Recent changes in extreme rainfall events in Peninsular Malaysia: 1971–2005. *Theor Appl Climatol* 99, 303–314.

ACCEPTED MANUSCRIPT

**List of Tables**

**Table 1** Geographic location of urban centres of the districts in Rajasthan State

**List of Figures**

**Fig. 1.** Location of urban centres considered in the State of Rajasthan, India.

**Fig. 2.** Location of urban centres with increasing, decreasing and no trends at 10% significance level for a) monsoon, b) non-monsoon and c) annual average rainfall.

**Fig. 3.** Location of urban centres with increasing, decreasing and no trends for a) monsoon, b) non-monsoon and c) annual average minimum temperature.

**Fig. 4.** Location of urban centres with increasing, decreasing and no trends at 10% significance level for a) monsoon, b) non-monsoon and c) annual average temperature.

**Fig. 5.** Location of urban centres with increasing, decreasing and no trends for a) monsoon, b) non-monsoon and c) annual average maximum temperature.

**Fig. 6.** Location of urban centres with increasing, decreasing and no trends at 10% significance level for extreme annual daily rainfall.

**Fig. 7.** Trends in statistically significant extreme annual daily rainfall in the urban centres of districts of Rajasthan State.

**Fig. 8.** Location of urban centres with increasing, decreasing and no trends at 10% significance level for extreme annual daily minimum temperature.

**Fig. 9.** Location of urban centres with increasing, decreasing and no trends at 10% significance level for extreme annual daily maximum temperature.

**Table 1** List of urban centers, size and population.

ID	Urban centre	Longitude	Latitude	Elevation (m.a.s.l.), m	Size (km <sup>2</sup> )** Approx.	Population (2011 census)
1	Ajmer	74° 42' E	26° 27'	486	31	542,580
2	Alwar	76° 38' E	27° 34'	271	-	315,310
3	Banswara	74° 24' E	23° 30'	302	-	100,128
4	Baran	76° 33' E	25° 05'	262	-	118,157
5	Barmer*	71° 25' E	25° 45'	227	-	83,517
6	Bharatpur	77° 30' E	27° 15'	183	-	252,109
7	Bhilwara	74° 40' E	25° 21'	421	-	360,009
8	Bikaner	73° 22' E	28° 01'	242	65	647,804
9	Bundi	75° 41' E	25° 27'	268	-	102,823
10	Chittorgarh	74° 42' E	24° 54'	394	-	116,409
11	Churu	75° 01' E	28° 19'	292	-	119,846
12	Dausa*	76° 20' E	26° 53'	333	-	61,589
13	Dholpur	77° 52' E	26° 41'	177	-	126,142
14	Dungarpur	73° 43' E	23° 50'	225	-	43000
15	Ganganagar	73° 50' E	29° 49'	164	-	224,773
16	Hanumangarh	74° 21' E	29° 35'	177	-	151,104
17	Jaipur	75° 52' E	26° 55'	431	168	3,073,350
18	Jaisalmer	70° 57' E	26° 55'	229	-	78,000
19	Jalore	72° 58' E	25° 22'	178	-	54185
20	Jhalawar*	76° 10' E	24° 35'	312	-	48,054
21	Jodhpur	73° 04' E	26° 18'	231	83	1,033,918
22	Jhunjhunun	75° 20' E	28° 06'	323	-	118,966
23	Karauli	77° 04' E	26° 30'	275	-	105,690
24	Kota	75° 52' E	25° 10'	271	62	1,001,365
25	Nagaur	73° 40' E	27° 00'	302	-	100,618
26	Pali	73° 25' E	25° 46'	214	-	229,956
27	Rajsamand*	73° 53' E	25° 04'	547	-	55,671
28	Sawai	76° 21' E	26° 01'	266	-	120,998
29	Sikar	75° 15' E	27° 36'	427	-	237,579
30	Sirohi*	72° 51' E	24° 53'	321	-	35,544
31	Tonk	75° 47' E	26° 10'	289	-	165,363
32	Udaipur	73° 42' E	24° 34'	598	-	451,735
33	Pratapgarh	74° 47' E	24° 02'	491	-	48,900

\*Population census 2001

\*\*For size of urban centre of district headquarter Rajasthan see e.g. Jhalawar, <https://maps.google.co.in/maps?q=Jhalawar&ie=UTF8&hq=&hnear=0x39653ede53cfaf07:0x31263af08bba421e,Jhalawar,+Rajasthan&gl=in&ei=TekFUuK8G4bLrQfk74F4&ved=0CLABELYD>. More details are also available in <http://www.citypopulation.de/php/india-rajasthan.php>

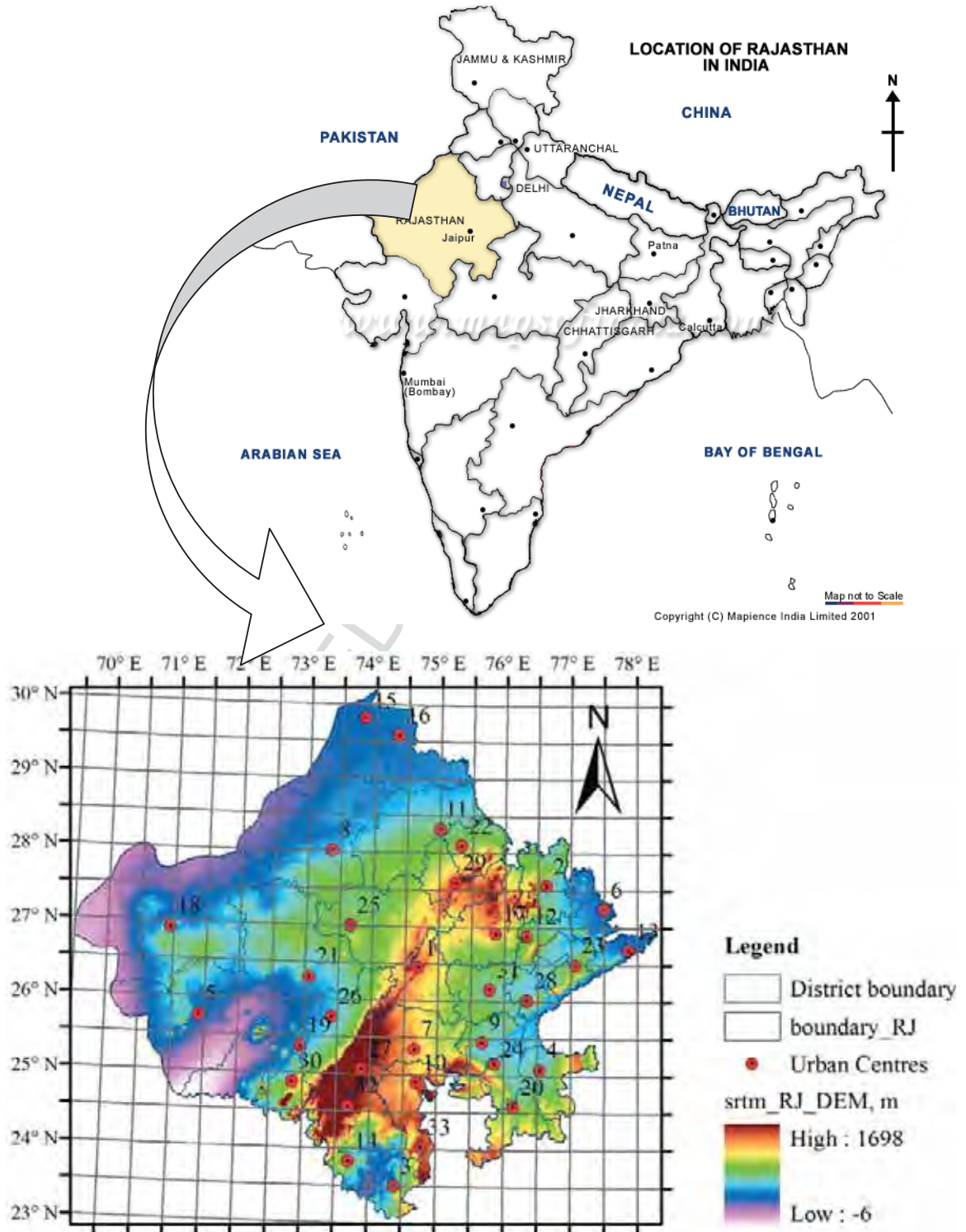
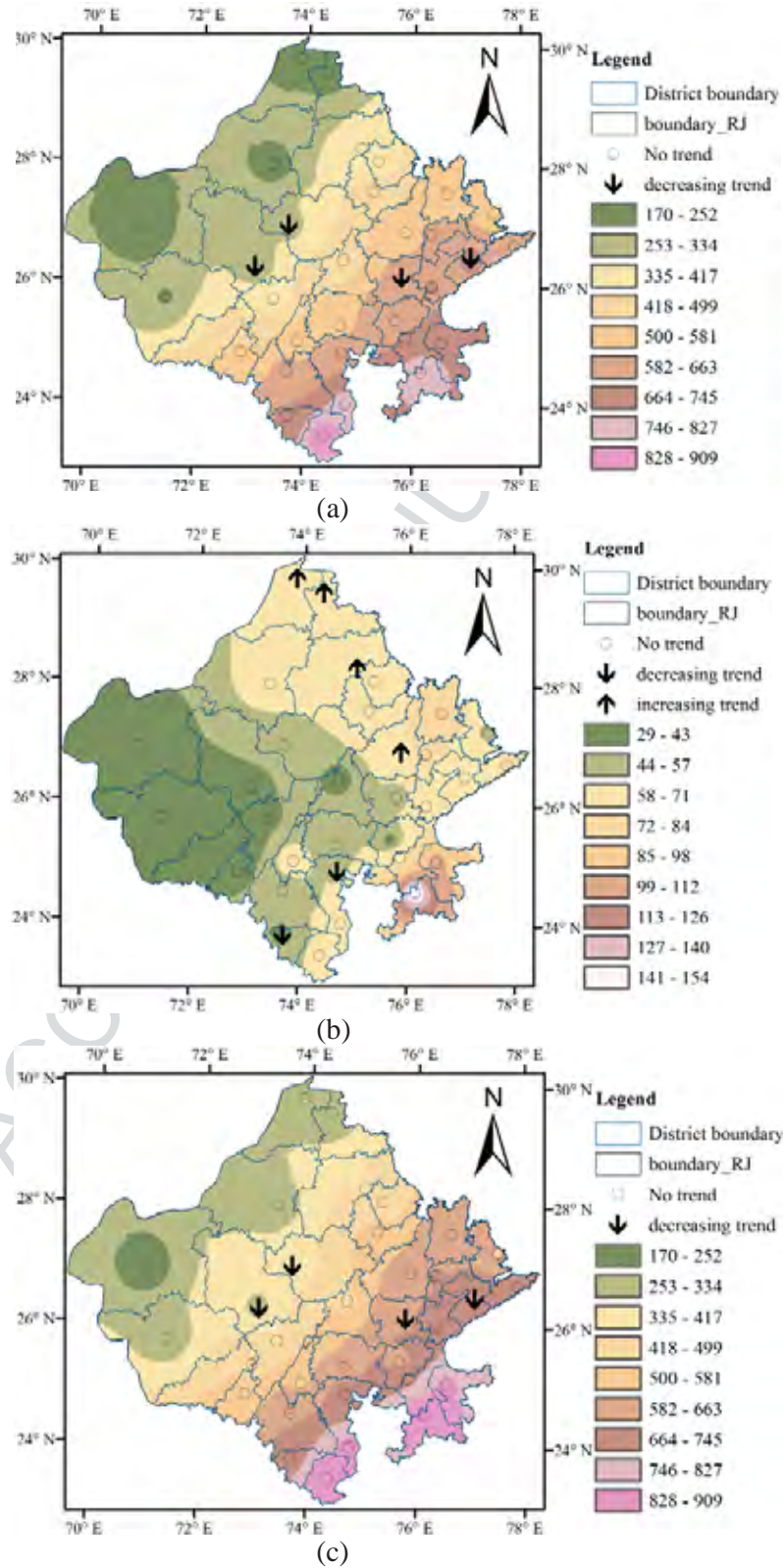
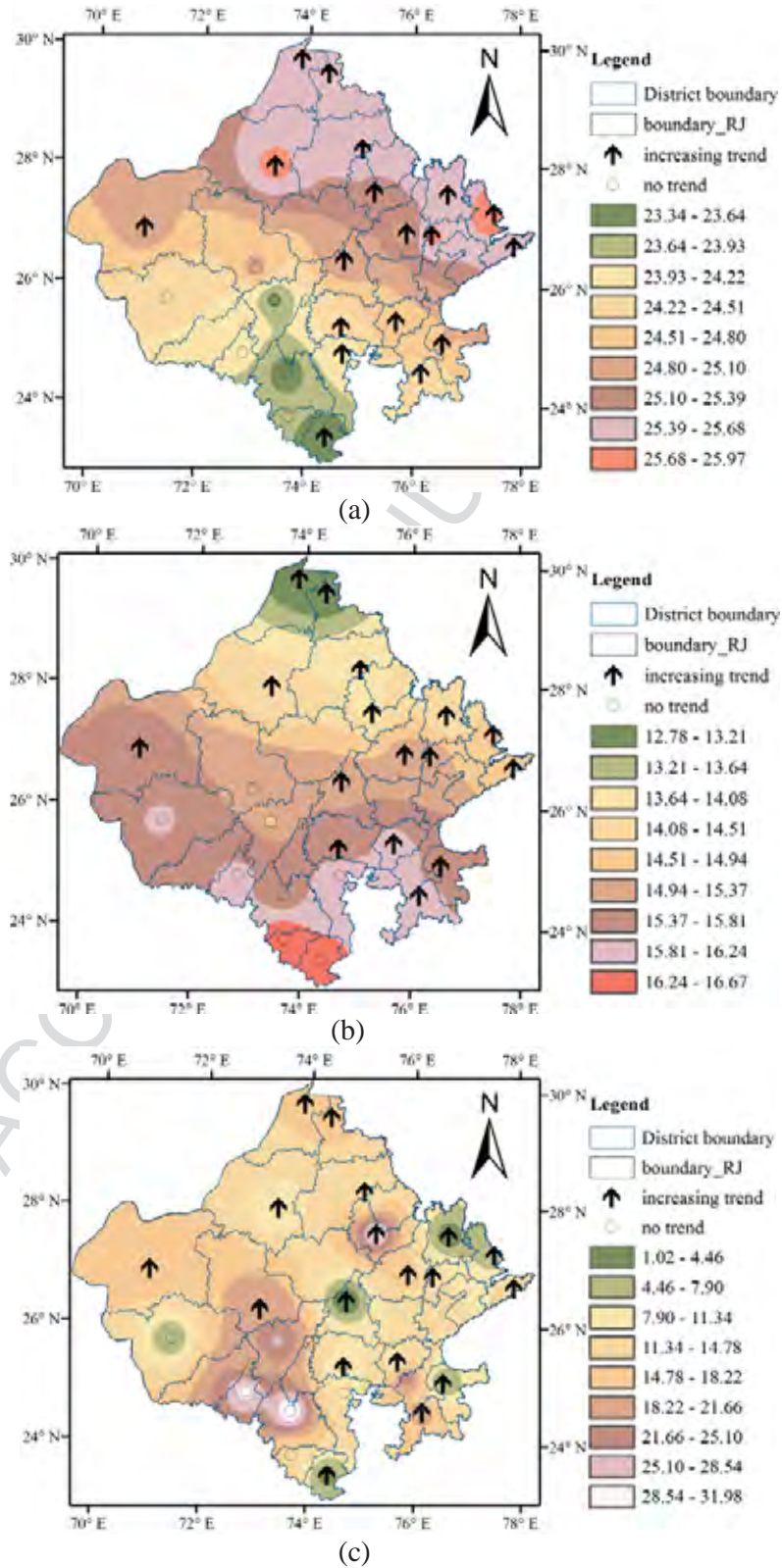


Fig. 1. Location of urban centres considered in the State of Rajasthan, India.

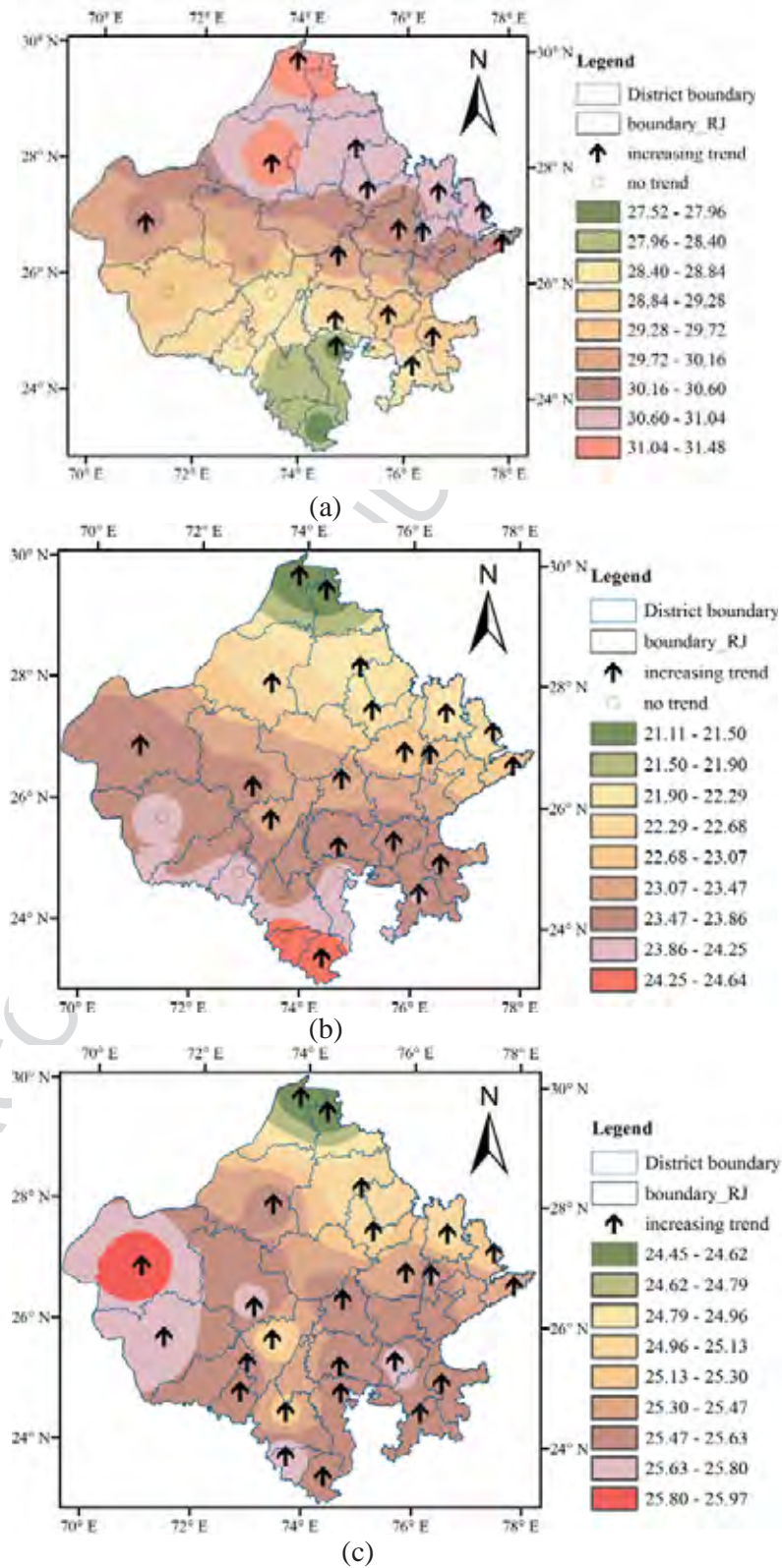




**Fig. 2.** Location of urban centres with increasing, decreasing and no trends at 10% significance level for a) monsoon, b) non-monsoon and c) annual average rainfall.

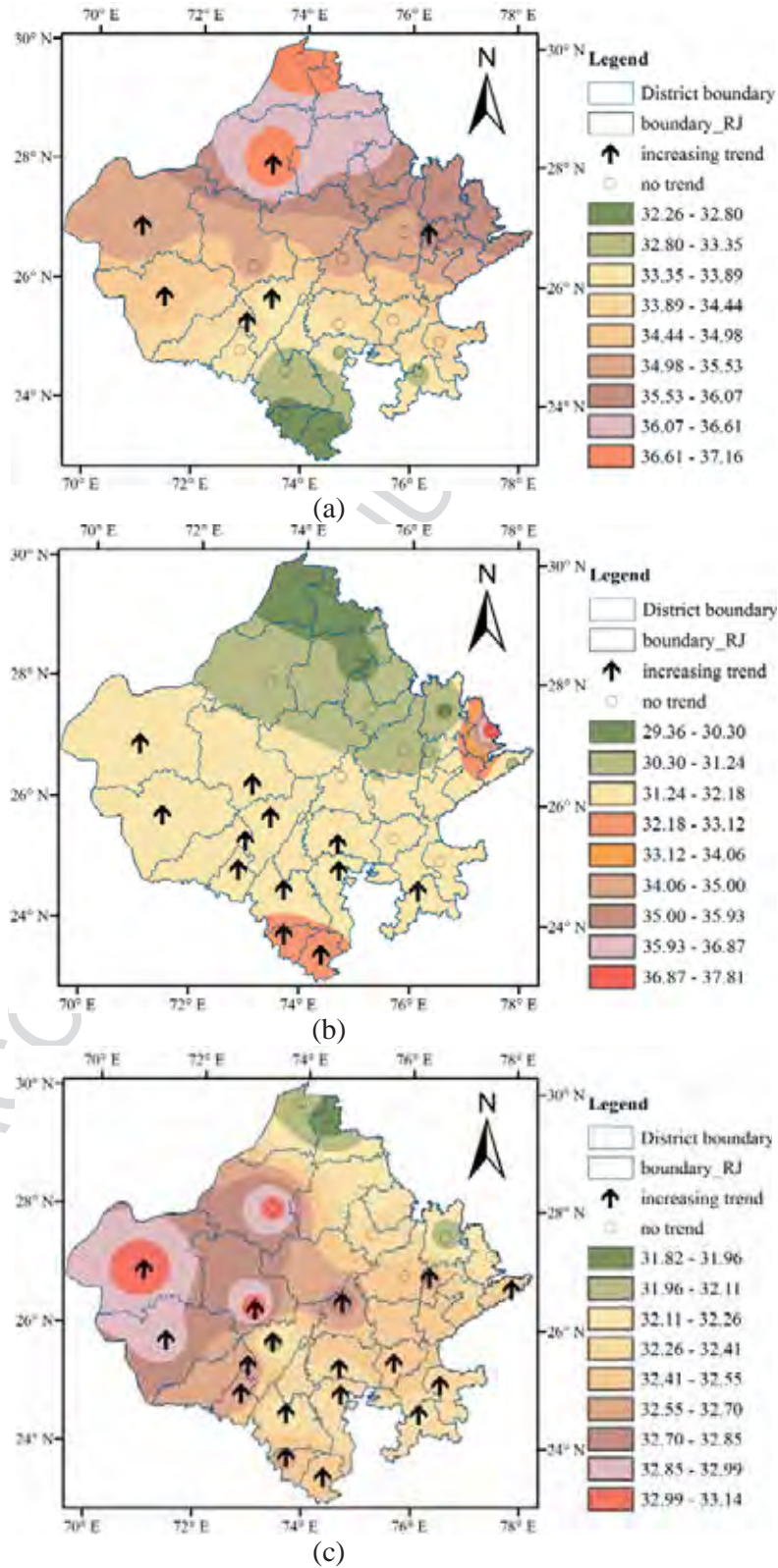


**Fig. 3.** Location of urban centres with increasing, decreasing and no trends for a) monsoon, b) non-monsoon and c) annual average minimum temperature.



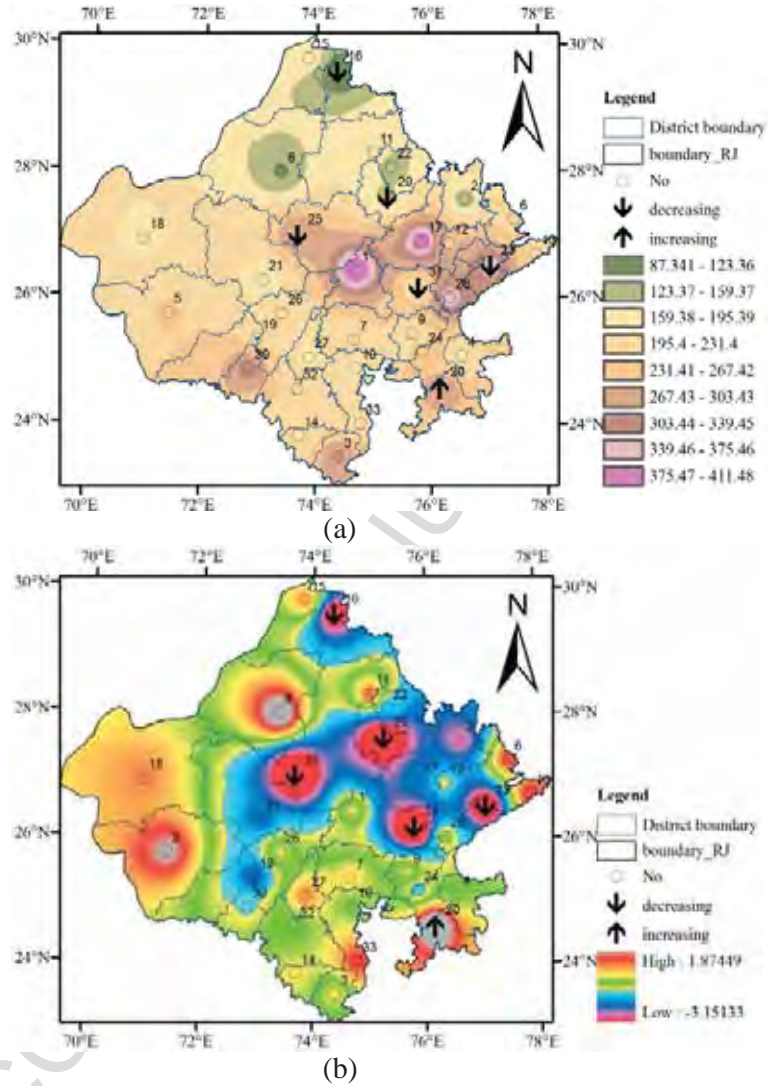
**Fig. 4.** Location of urban centres with increasing, decreasing and no trends at 10% significance level for a) monsoon, b) non-monsoon and c) annual average temperature.



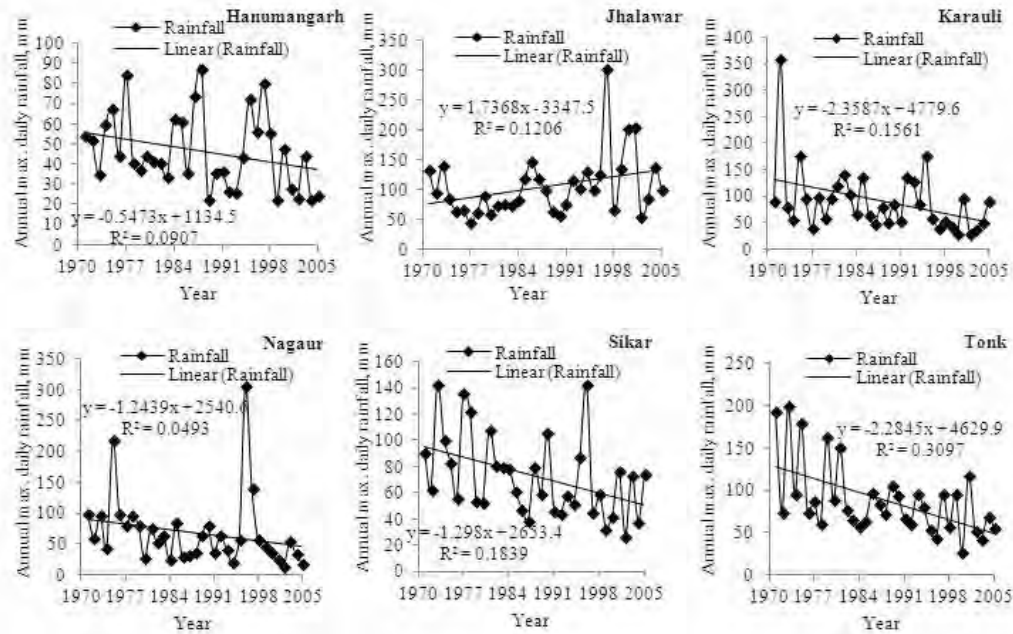


**Fig. 5.** Location of urban centres with increasing, decreasing and no trends for a) monsoon, b) non-monsoon and c) annual average maximum temperature.

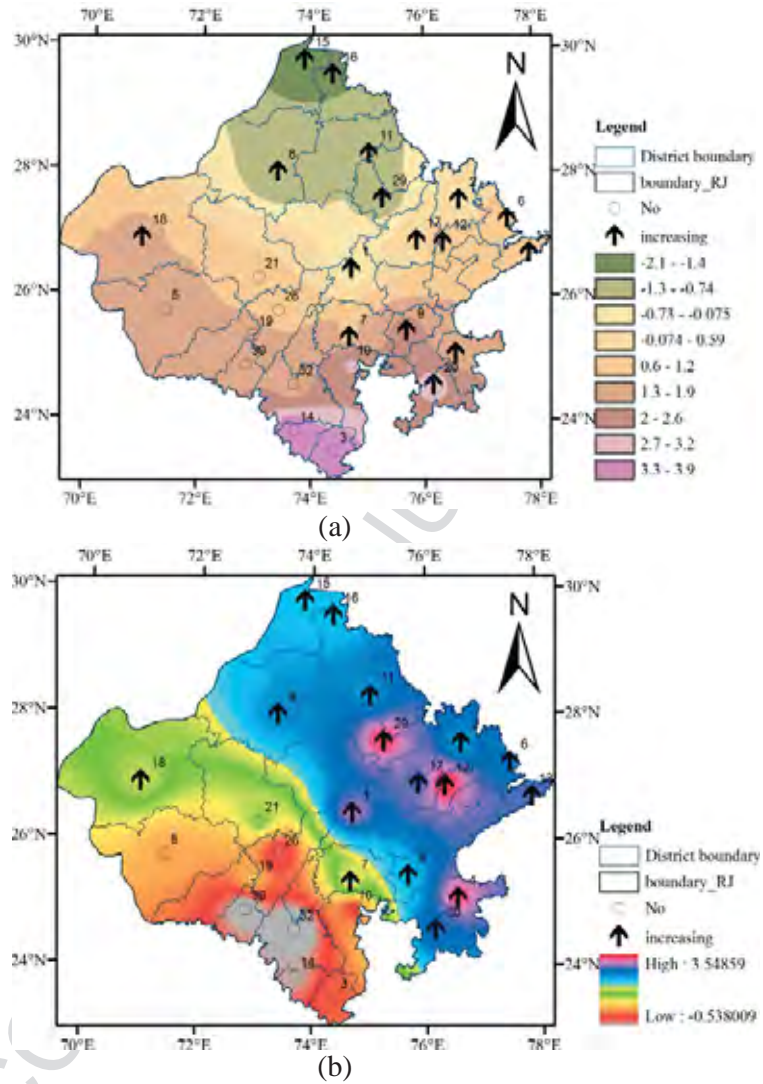




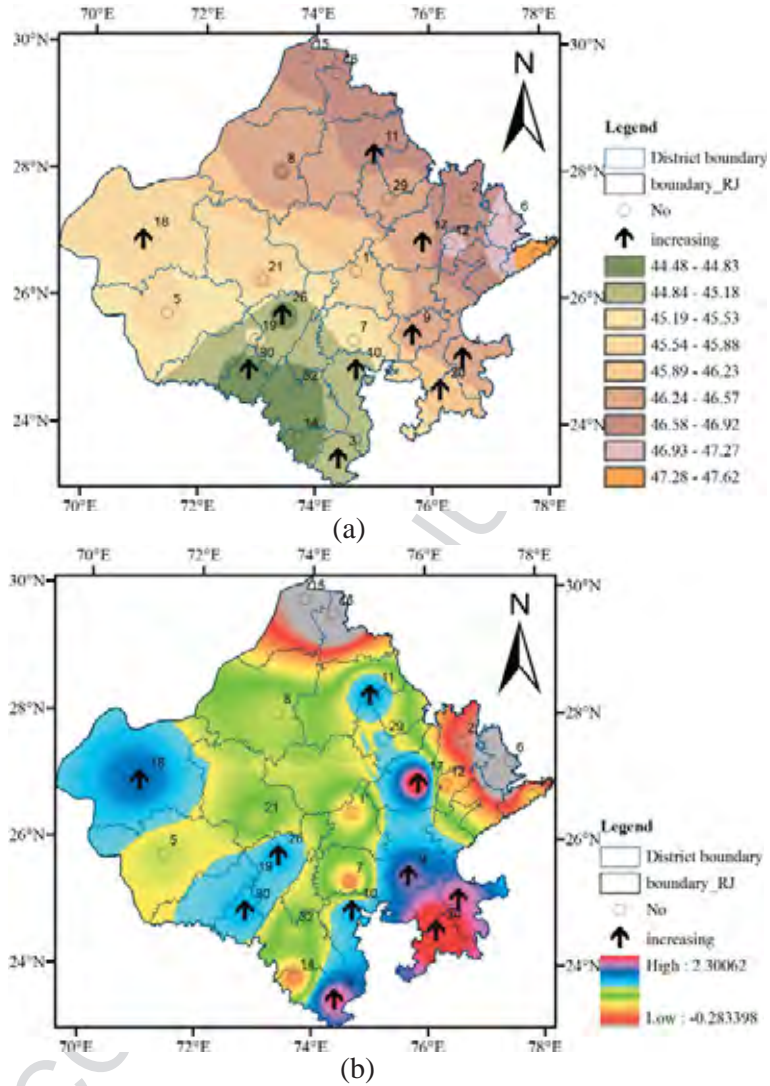
**Fig. 6.** Location of urban centres with increasing, decreasing and no trends at 10% significance level for extreme annual daily rainfall.



**Fig. 7.** Trends in statistically significant extreme annual daily rainfall in the urban centres of districts of Rajasthan State.



**Fig. 8.** Location of urban centres with increasing, decreasing and no trends at 10% significance level for extreme annual daily minimum temperature.



**Fig. 9.** Location of urban centres with increasing, decreasing and no trends at 10% significance level for extreme annual daily maximum temperature.



**Highlights**

- Highlights the importance of climate change assessments to prepare suitable adaptation strategies in urban centres in India
- Identifies trends in mean (annual and seasonal) and extreme annual daily rainfall and temperature in the 33 urban centres of Rajasthan State
- Found predominant changes in mean and extreme rainfall and temperature trends in most of the 33 urban centres of Rajasthan State

ACCEPTED MANUSCRIPT

## A Spectral Analysis Based Methodology to Detect Climatological Influences on Daily Urban Water Demand

Jan Adamowski · Kaz Adamowski ·  
Andreas Prokoph

Received: 2 April 2012 / Accepted: 3 October 2012 / Published online: 19 October 2012  
© International Association for Mathematical Geosciences 2012

**Abstract** Urban water demand (UWD) is highly dependent on interacting natural and socio-economic factors, and thus a wide range of data analysis and forecasting methods are required to fully understand the issue. This study applies, for the first time, the continuous wavelet transform to determine changes in the temporal pattern of UWD and its potential meteorological drivers for three major Canadian cities: Calgary, Montreal, and Ottawa. This analysis is complemented by Fourier and cross-spectral analysis to determine inter-relationships and the significance of the patterns detected. The results show that the annual (365 days) cycle provides the most consistent and significant relationship between UWD and meteorological drivers. Wavelet analysis shows that UWD is only sensitive to air temperature in the summer months when mean daily temperatures are greater than 10 to 12 °C. For the three cities studied, the UWD increases by between 10 ML (Montreal) and 50 ML (Calgary) per day with every 1 °C increase in air temperature. In an area with low precipitation (Calgary), there is an inverse relationship between UWD and precipitation during summer months. Wavelet transform and Fourier analysis also detected a 7-day cycle in UWD, particularly in the more industrialized city of Montreal, which is related to the working week. In general, applying the season dependent linear relationships between UWD and temperature is suggested as perhaps being more appropriate and potentially successful for forecasting, rather than continuous complex nonlinear algorithms that are designed to explain variability in the entire UWD record.

---

J. Adamowski  
Department of Bioresource Engineering, McGill University, 21 111 Lakeshore Road, Ste. Anne de  
Bellevue, QC, H9X 3V9, Canada

K. Adamowski  
Department of Civil Engineering, University of Ottawa, Ottawa, ON, K1N 6N5, Canada

A. Prokoph (✉)  
SPEEDSTAT, 19 Langstrom Crescent, Ottawa, ON, K1G 5J5, Canada  
e-mail: [aprokocon@aol.com](mailto:aprokocon@aol.com)

**Keywords** Precipitation · Air temperature · Continuous wavelet transform · Fourier analysis

## 1 Introduction

The need for defining the factors that influence urban water demand (UWD) has dramatically increased in recent decades as population growth, industrialization, and an increased likelihood of extreme climate events due to anthropogenic global warming lead to challenges in balancing water supply and demand (IPCC 2007). Although Canada holds approximately 7 % of the world's freshwater and only 0.5 % of the world's population (Environment Canada 2010), many of Canada's cities are experiencing some degree of water stress. This is due to a complex combination of factors, including population growth and water losses from ageing infrastructure, as well as the fact that most of the freshwater flows north away from the main areas of population (Adamowski et al. 2009). In addition, Canada is also experiencing the regional impact of global climate change, which makes future levels of supply and demand uncertain. For example, Zhang et al. (2000) highlighted that annual precipitation totals have generally increased by 5 % to 35 % across Canada between 1950 and 1998, although the most significant increases occurred north of 60° N. Despite often limited water availability and uncertain future supplies, few studies on detecting relationships between temporal patterns in UWD and meteorological records in Canada have been carried out to date (Adamowski and Karapataki 2008; Adamowski et al. 2012). While meteorological records in major Canadian cities are mostly over 50 years long, UWD records are much shorter and often incomplete. Only three cities (Calgary, Ottawa, and Montreal) in Canada have readily available recorded daily urban water demand and meteorological data simultaneously without gaps for intervals of more than five successive days. These three cities represent different climatic and economic environments, with Calgary having a continental dry climate and minor manufacturing, Ottawa having a moderately continental humid climate and minor manufacturing, and Montreal having a moderately continental humid climate and strong manufacturing industry.

Research into UWD has increased since the end of the twentieth century, with particular focus on forecasting UWD and describing the complexity of the factors that influence it (House-Peters and Chang 2011). Such factors are highly complex, but can be divided into socio-economic factors and natural factors. Socio-economic variables such as water price, rate structures, and population growth exhibit mostly nonlinear temporal patterns and are analyzed accordingly (Arbués et al. 2004, 2010). Research on the influence of natural factors on UWD has focused especially on meteorological factors that are incorporated in complex log-linear water demand models (Miaou 1990; Schleich and Hillenbrand 2009). Traditional statistical models of urban or residential demand have tended to be based on a wide range of socio-economic and natural variables, including meteorological variables such as temperature, precipitation, days of sunshine, and wind speed. Commonly used models include multiple linear regression (MLR) and auto-regressive moving average (ARIMA), which acknowledge seasonal variability in both water supply and urban demand (Anderson et al. 1980;

Maidment and Parzen 1984). For example, Agthe and Billings (1980) included a time lagged value in a multiple regression model to better account for the influence of past water use on current water use. Maidment et al. (1985) used regression analysis to model the association between daily municipal water consumption and rainfall and air temperature.

Several studies acknowledged the different UWD dependencies on seasonal sensitive (summer) weather compared to relatively non-sensitive (winter) weather by implementing a sinusoidal variability model (Gato et al. 2007; Praskievicz and Chang 2009). In addition, Zhou et al. (2000) developed a method that recognizes that seasonal variations in water consumption are not completely the result of cyclic patterns of air temperature and evaporation over a year, but are also related to abrupt and extreme events, such as floods and drought. Moreover, short-time extreme precipitation events that lead to rapid UWD reductions cannot be reconstructed or modeled using sinusoidal models or Fourier analysis. Thus, auto-regressive methods with short-term memory functions have been developed to address the relationships between meteorological events and UWD changes (Agthe and Billings 1980; Zhou et al. 2000).

Maidment and Miaou (1986) demonstrated that models assuming linear relationships between weather variables and water demand are not adequate. For example, it has been suggested that water use responds to the occurrence of rainfall rather than the amount that falls (Martínez-Espiñeira 2002). Furthermore, it is likely that there are thresholds in temperature and/or precipitation, above which water demand changes. For example, Gato et al. (2007) identified thresholds that separate the base water use, which is independent of temperature and rainfall, from seasonal water use. A further drawback of many of the statistical models of UWD and meteorological factors described above is that they are based on an assumption of stationarity in data.

The wavelet transform (WT) is an alternative method that has not previously been used for detecting relationships between temporal patterns in UWD and meteorological records. In contrast to MLR, ARIMA, ANN, or Fourier analysis used for UWD and meteorological records (House-Peters and Chang 2011, and references therein), WT permits the detection of temporal patterns synchronously according to their scale (or frequency/wavelength) at varying temporal resolutions (Rioul and Vetterli 1991). Therefore, signals can be extracted according to their scale as well as their time of occurrence, meaning for example that UWD cycles restricted to a certain season, as well as short-term stochastic events such as precipitation, can be identified. Continuous WT (CWT) permits signal detection at a very wide range of bandwidth and temporal resolutions, in contrast to the more commonly used discrete WT (DWT) (Adamowski et al. 2012). The multi-resolution feature of the CWT leads to higher precision in defining intervals and wavelengths of significant UWD variability than Fourier analysis. Fourier and cross-spectral analysis can be used to support WT results due to their more sophisticated definitions of significance and confidence, as was done in this study.

This study aims to demonstrate the relevance of a spectral analysis based methodology that includes the continuous wavelet transform for determining temporal-frequency dependency patterns. A particular goal of this study is to compare



differences in urban water demand and meteorological/climatological factors in different climatic and economic regions of Canada.

## 2 Methodology

The methodology consists of:

1. The wavelet transform is used to qualitatively detect periodic sinusoidal signals according to their time intervals of occurrence, as well as their wavelengths in meteorological and corresponding UWD records for the three selected major urban centers in Canada.
2. Time intervals of the records exhibiting different cyclicity and/or variability patterns that were detected by the wavelet transform are separately analyzed for links between meteorological records and UWD records.
3. Fourier analysis is then used to quantify the significance of the detected cycles.
4. Cross-spectral analysis is applied to quantify the correlation between the time series at specific wavelengths between meteorological and UWD records.

The steps of analysis and the results are described for each location separately, in order to emphasize the procedure and location specific features. The results are then discussed in relation to specific locations.

## 3 Time Series Analysis Background

### 3.1 Wavelet Analysis

Wavelet analysis was first developed as a filtering and data compression method for geophysical exploration in the 1980s (Morlet et al. 1982). It transforms a time series into a frequency domain by simultaneously transforming the depth or time domain and the scale or frequency domain using various shapes and sizes of short filtering functions called wavelets. The CWT method allows for the automatic localization of periodic signals, gradual shifts, abrupt interruptions, trends, and onsets of trends in time series data (Rioul and Vetterli 1991). The wavelet coefficients  $W$  of a time series  $x(s)$  are calculated by a simple equation

$$W_{\psi}(a, b) = \left( \frac{1}{\sqrt{a}} \right) \int x(s) \psi \left( \frac{s-b}{a} \right) ds, \quad (1)$$

where  $\psi$  is the mother wavelet,  $a$  is the scale factor that determines the characteristic frequency or wavelength, and  $b$  represents the shift of the wavelet over  $x(s)$  (Prokoph and Barthelmes 1996).

The bandwidth resolution for a wavelet transform varies with  $\Delta a = \frac{\sqrt{2}}{4\pi al}$ , and a location resolution  $\Delta b = \frac{al}{\sqrt{2}}$ . Due to Heisenberg's uncertainty principle  $\Delta a \Delta b \geq 1/4\pi$ , the resolution of  $\Delta b$  and  $\Delta a$  cannot be arbitrarily small (Prokoph and Barthelmes 1996). Parameter  $l$  is used to modify the wavelet transform bandwidth resolution either in favor of time or in favor of frequency.

In this study, the CWT was used with the Morlet wavelet as the mother function (Morlet et al. 1982), which is expressed as

$$\psi(s) = \pi^{-\frac{1}{4}} e^{-i2\pi f_0 s} e^{-\frac{s^2}{2}}. \quad (2)$$

The Morlet wavelet is a sinusoid with wavelength/scale  $a$  modulated by a Gaussian function (Torrence and Compo 1998; Adamowski et al. 2009). The Morlet wavelet has been widely and successfully used on hydrological and meteorological records (Ware and Thomson 2000; Coulibaly 2006). The matrix of the wavelet coefficients  $W_l(a, b)$ , the so-called ‘scalogram’, was coded in a color scale (orange highest, blue lowest  $W_l(a, b)$ ), for better graphical interpretation. Edge effects of the wavelet coefficients occur at the beginning and end of the analyzed time-series and increase with increasing wavelength (scale) and parameter  $l$ . Thus, a ‘cone of influence of edge effects’ is formed (Torrence and Compo 1998). The cones of edge effect influences of greater than 10 % of bias due to non-available data points in the analysis windows behind the edges are illustrated in the scalograms.

The wavelet coefficients  $W$  are normalized by using the L1 normalization ( $1/a$ ), replacing the commonly used  $1/\sqrt{a}$  L2 or L<sup>2</sup> normalization (see Eq. 1). This allows wavelet coefficients to be interpreted in terms of Fourier amplitudes (Prokoph and Barthelmes 1996). In addition, the L2 normalization of the Morlet wavelet commonly leads to overvaluing wavelet coefficients in long wavelengths compared to shorter ones, as discussed in detail by Schaeffli et al. (2007). Detailed explanations of the advantages and disadvantages of the normalization types, in terms of accuracy of the energy spectrum, amplitudes and white noise, variance and bias of arbitrary estimated continuous wavelet spectra depending on the algorithms applied are provided by Maraun and Kurths (2004) and Maraun et al. (2007). The parameter  $l = 6$  was chosen for all analyses as it provides sufficiently precise results in the resolution of time and scale, and is commonly suggested for hydrological and meteorological records (Schaeffli et al. 2007; Adamowski et al. 2009). Significance tests for the WT were not included in this study since the significance tests for WT, as for example suggested by Torrence and Compo (1998), are based on the assumption of combined white and red-noise for windows of discrete scales (frequency)–location (time-intervals), but have not been proven to correctly evaluate the significance of signals of deterministically changing frequencies and amplitude (so-called “chirps”). For this reason, this study applied significance tests on Fourier analysis for time-intervals proven (by WT) to have approximately stationary signals since this was deemed to be better suited for this research.

The wavelet analysis technique used in this article is explained in detail in Prokoph and Barthelmes (1996) and Adamowski et al. (2009).

### 3.2 Fourier Analysis

Fourier transform (Fourier analysis) is defined by

$$P_f = \int x(t) e^{i2\pi f t} dt \quad t = 1 \dots N, \quad (3)$$

with  $x(t)$  as the discrete time series,  $f$  the frequency, and  $P$  as the spectral power (Davis 2002). The white noise level  $P_w$  and the red noise level were calculated

from the lag-one auto-correlation coefficient  $r$  according to an auto-regressive model (Mann and Lees 1996), using a  $\chi^2$  value for the chosen confidence level (95 %) at 1 degree of freedom. The value of  $B_f$  is the Fourier power of the background spectrum for the Fourier frequencies  $k$  combining the red and white noise spectrum and the confidence level set by the  $\chi^2$  value. The Fourier power  $P_w$  of a white noise spectrum is equal to the average power of all frequencies  $f$

$$B_f = P_w \chi^2 \frac{1 - r^2}{1 + r^2 - 2r \cos(2\pi f/N)}. \quad (4)$$

Only frequencies (or wavelengths) where  $P_f^2 > B_f$  for confidence levels greater than 95 % were considered to be significant in this study. Fourier analysis and confidence level calculations were carried out using the REDFIT software (Schulz and Mudelsee 2002).

### 3.3 Cross-spectral Analysis

Cross-spectral analysis is an extension of Fourier analysis for two time series (Bassett and Tinline 1970). The cross spectrum  $S_{yz}$  of two time-series  $y$  and  $z$  with power spectra  $S_{yy}$  and  $S_{zz}$ , respectively, is defined by

$$S_{yz}(f) = dS_{yy}(2\pi f) dS_{zz}^*(2\pi f). \quad (5)$$

It should be noted that the cross spectrum is a complex variable with a real and an imaginary component

$$S_{yz}(f) = Co(2\pi f) + iQuad(2\pi f), \quad (6)$$

where  $Co$  is the co-spectrum (in-phase covariance), which constitutes the real component, and  $Quad$  is the quadrature spectrum (out-off-phase covariance), which constitutes the imaginary component. The cross spectrum can be normalized to obtain the spectral equivalent of the correlation coefficient, the coherency spectrum

$$R_{yz}(f) = \frac{|S_{yz}(f)|}{S_{yy}(f)S_{zz}(f)}. \quad (7)$$

The coherency spectrum can be tested for confidence in a similar way to the correlation coefficient. As for the spectral estimate, the numerical solution reflects that the time-series is finite using  $N$  equally spaced locations. The cross spectrum provides the density of covariance between two variables ( $y, z$ ) in the Fourier domain (the frequency of the sine and cosine wave domain)

$$S_{yz}(fk) = \frac{1}{M} \sum_{i=1}^M \left\{ \frac{2\Delta x}{(2m+1)N^2} \sum_{l=-m}^m C_{iyz}(fk+l) + iQ_{iyz}(fk+l) \right\}, \quad (8)$$

with

$$C_{iyz}(fk) = A_{iy}A_{iz} + B_{iy}B_{iz}, \quad (9)$$

and

$$Q_{iyz}(fk) = A_{iz}B_{iy} - A_{iy}B_{iz}, \quad (10)$$

**Fig. 1** Map of Canada with survey locations

with the real component (*co-spectrum*) for  $M$  averaged transects

$$C_{yz}(f_k) = \frac{1}{M} \sum_{i=1}^M \left\{ \frac{2\Delta x}{(2m+1)N^2} \sum_{l=-m}^m C_{iyz}(f_{k+l}) \right\}, \quad (11)$$

and the imaginary component (*quadrature spectrum*)

$$Q_{yy}(f_k) = \frac{1}{M} \sum_{i=1}^M \left\{ \frac{2\Delta x}{(2m+1)N^2} \sum_{l=-m}^m Q_{iyz}(f_{k+l}) \right\}. \quad (12)$$

The squared coherency spectrum is calculated by

$$R_{yz}^2(f) = \frac{C^2 + Q^2}{S_{yy}(f)S_{zz}(f)} = \frac{|S_{yz}(f)|^2}{S_{yy}(f)S_{zz}(f)}. \quad (13)$$

Here, we used a 3-spectral-estimate window for smoothing (Patterson et al. 2004). According to Jenkins and Watts (1968), the squared coherency can be tested for two variables by using the  $F$ -test:

$$F_{crit} = \frac{R_{yz}^2(f)}{1 - R_{yz}^2(f)} \left( \frac{(2M(2m+1) - 4)}{4} \right). \quad (14)$$

#### 4 Data and Study Site

The proposed WT approach was tested on three large Canadian cities, Calgary, Montreal, and Ottawa (Fig. 1) that each has more than 700,000 inhabitants, but have different climatic and socio-economic characteristics.

Calgary has a continental climate characterized by cold winters, occasionally interrupted by milder phases, and warm summers. Its climate is generally dry, although it experiences intermittent high precipitation events, such as thunderstorms or blizzards (Environment Canada 2012). Calgary is a major administrative center of the oil industry, but has no significant manufacturing industry. Residential water use is



approximately 257 liters per person per day and accounts for 52 % of the city's water consumption (City of Calgary 2012).

Both Ottawa and Montreal are located in the Saint Lawrence lowlands, which have a moderate continental climate with cold winters and humid, hot summers (Environment Canada 2012). In Ottawa, water consumption was 291 liters per person per day in 2009, which represents a decrease of 9 % since 2007, following implementation of the city's Water Efficiency Strategy (City of Ottawa 2010). The City of Montreal supplies drinking water to 1.8 million people and produces 1300 liters per person per day (Aubertin et al. 2002). The almost five times higher supply in Montreal compared to Ottawa and Calgary is due to a combination of industrial demand and securing supply for extreme peak periods. In addition, approximately 40 % of Montreal's water supply is lost in the piping due to the poor condition of the water supply piping infrastructure in the city of Montreal; this also contributes to the much higher supply that is produced in Montreal compared to Ottawa.

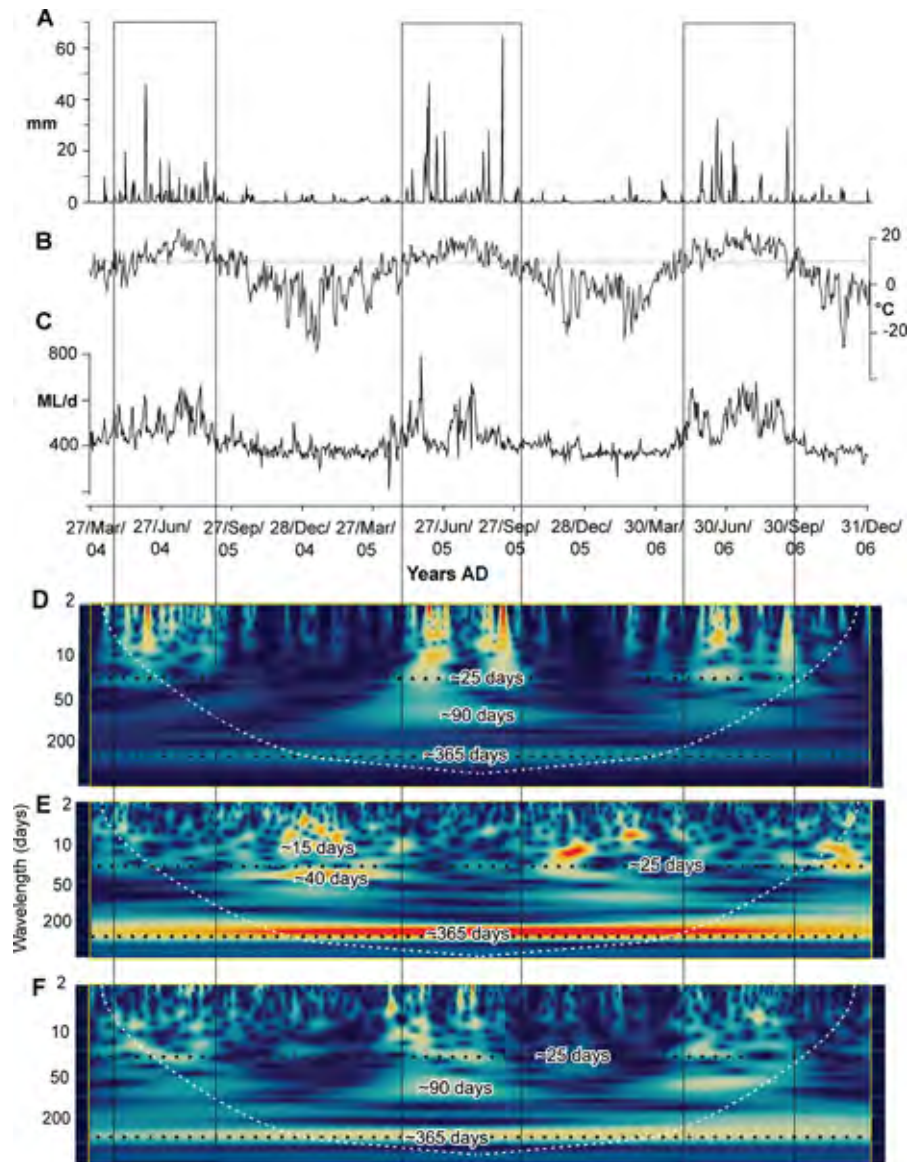
Daily UWD data (consumption in ML/day) for the entire municipal areas of Montreal, Ottawa and Calgary were recorded discontinuously for the last  $\sim 50$  years for each city. For this study, a string of continuous and almost complete UWD and meteorological records of at least 1,000 days were used for each location. For the same period, mean daily air temperature and total precipitation data for the city centers were obtained from Environment Canada (Environment Canada 2012). The records include very few missing data points for less than five successive days. The gaps that did exist were addressed via linear interpolation.

For the Fourier and cross-spectral analysis, a 730-day subset (January 1st to December 31st of the following year) of the data from each location were analyzed to allow for maximum compatibility of the significance test for the same frequencies in UWD and meteorological records.

## 5 Results

The UWD and meteorological records for Calgary ranged from March 2004 to December 2006, thus encompassing three complete summer seasons and two complete winter seasons (Fig. 2). During all three summer seasons, mean air temperatures were greater than 10 °C (dotted line in Fig. 2B), and the UWD was often greater than 600 ML/day (Fig. 2C), particularly during successive weeks of daily precipitation that were smaller than 2 mm/day as in the summer of 2006 (Fig. 2A). The most significant variations in UWD occurred during the summer and coincided with large precipitation events. In contrast, both the data (Figs. 2A–2C) and the wavelet analysis (Figs. 2D–2F) showed that the highest temperature variability occurred during winter, while the mean summer temperatures were relatively stable at around 20 °C. All records formed an annual ( $\sim 365$  day) cycle, with the highest temperature, precipitation and UWD values occurring during the summer months.

The WT analysis highlighted cycles of approximately 25 days during the summer season (three cycles per season) in both UWD and precipitation, indicating that rainfall waves reduced the UWD, and dry periods led to increases in UWD (Figs. 2A and 2C). However, the air temperature record did not show any consistent cyclical



**Fig. 2** Wavelet analysis of daily UWD and meteorological data from Calgary from 3/2004 to 12/2006: **A** daily total precipitation (mm); **B** daily mean temperature ( $^{\circ}\text{C}$ ) with sensitivity threshold (*dashed line*); **C** urban water demand (ML/d); **D** wavelet scalogram of daily total precipitation record; **E** wavelet scalogram of daily mean temperature record; **F** wavelet scalogram of daily urban water demand. *Curved dashed lines* mark “cone of influence” (see text), *horizontal dotted lines* mark 25-days scale, and *boxes* mark high UWD variability intervals. *Orange colors* indicate large wavelet coefficients and *blue colors* weak wavelet coefficients

pattern besides the annual cycle, except intermittent but strong fluctuations in a 14–40-day range during the winter and spring (Fig. 2E).

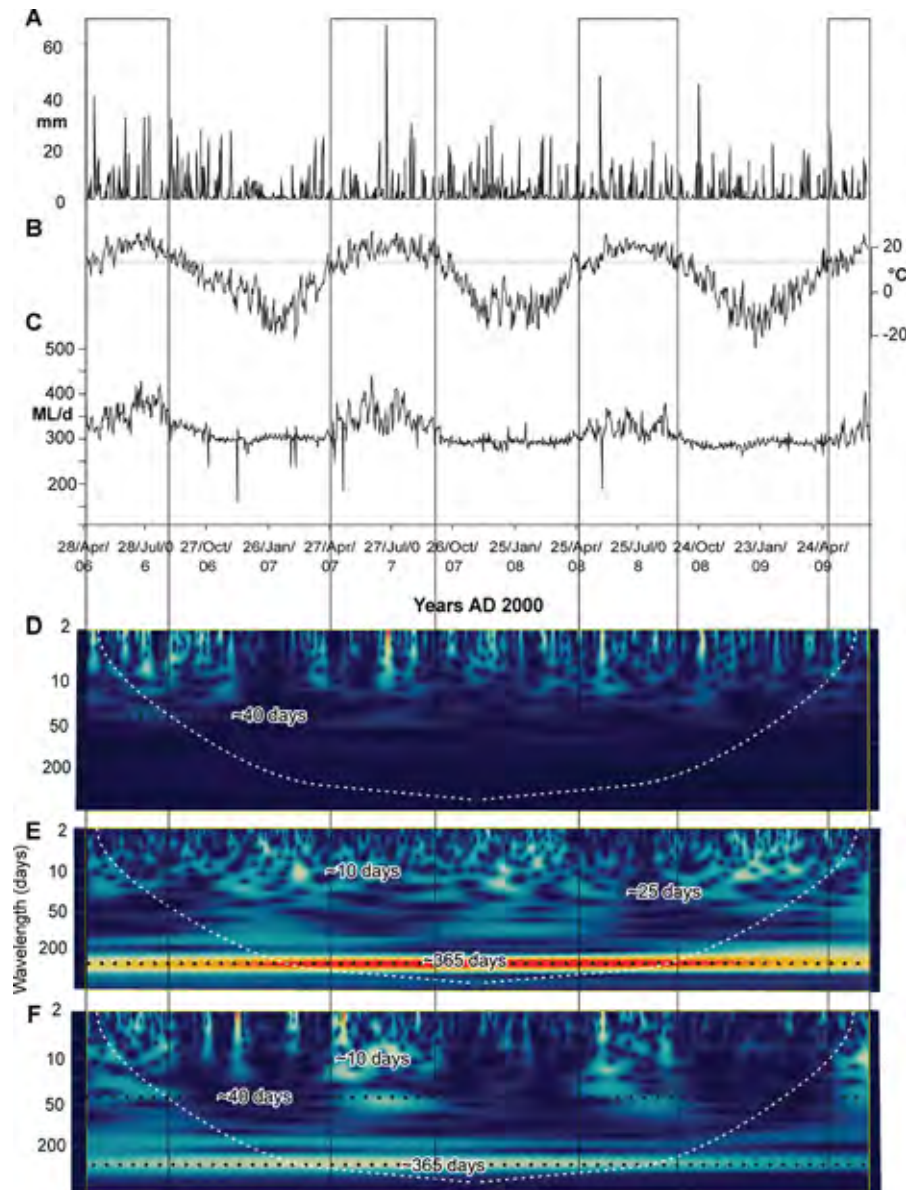
The UWD and meteorological records for Ottawa ranged from April 2006 to June 2009, spanning three complete summer and winter seasons (Fig. 3). During the summer period (May to September), the UWD was mostly greater than 350 ML/day, except after occasional periods of heavy rain (Figs. 3A and 3C). Periods of higher variability in UWD correlated with mean air temperatures of greater than 12 °C (Figs. 3E and 3F). In contrast to Calgary, the precipitation was much more evenly distributed through the year, and periods of more than two weeks during which there was no or little precipitation were almost absent (Fig. 3A). WT showed that the annual cycle (~ 365 days) dominated the air temperature and UWD pattern, superimposed by weaker 10–40 day cycles, predominately during the summer months.

In Montreal, the UWD and meteorological records ranged from February 1999 to June 2002 and encompassed three complete summer seasons and three complete winter seasons. The UWD was highest between May and September when mean daily temperature was mostly above 12 °C (Figs. 4B and 4C). Drought periods, as indicated by “white” (no signal) areas in the high-frequency field (< 10 days wavelength), occurred occasionally during the summer (Fig. 4D) and were associated with increased UWD. The UWD and air temperature records were dominated by the annual cycle (Fig. 4). This is a similar pattern to that which occurred in Ottawa, which is only 180 km from Montreal and has a similar climate. In addition, the UWD and total precipitation both have 180-day and ~ 50-day cycles in common, particularly during the summer months (Figs. 4D and 4F). WT also showed that the UWD is dominated by 7-day cycles related to the working week. However, during the summer this cycle was offset and vanished if large precipitation fluctuations occurred (Figs. 4A, 4C, and 4F).

Scatter plots for all three cities showed that there was no obvious relationship between daily UWD and mean daily temperature below a mean daily temperature of approximately 10 °C for Calgary, and approximately 12 °C for Ottawa and Montreal (Figs. 5A–5C). Above these temperature thresholds, the linear correlation between UWD and mean daily temperature became more than 99 % significant ( $p < 0.001$ ) (Table 1). This confirms the pattern depicted by the wavelet analysis (Figs. 2, 3, and 4). Table 2 shows that the variability in the UWD pattern exponentially increases to temperature in all cities when the temperature thresholds of 12 °C, 10 °C, and 12 °C for Ottawa, Calgary, and Montreal are crossed, respectively. In contrast, the precipitation variability increases approximately linearly with mean daily temperature increases (Table 2).

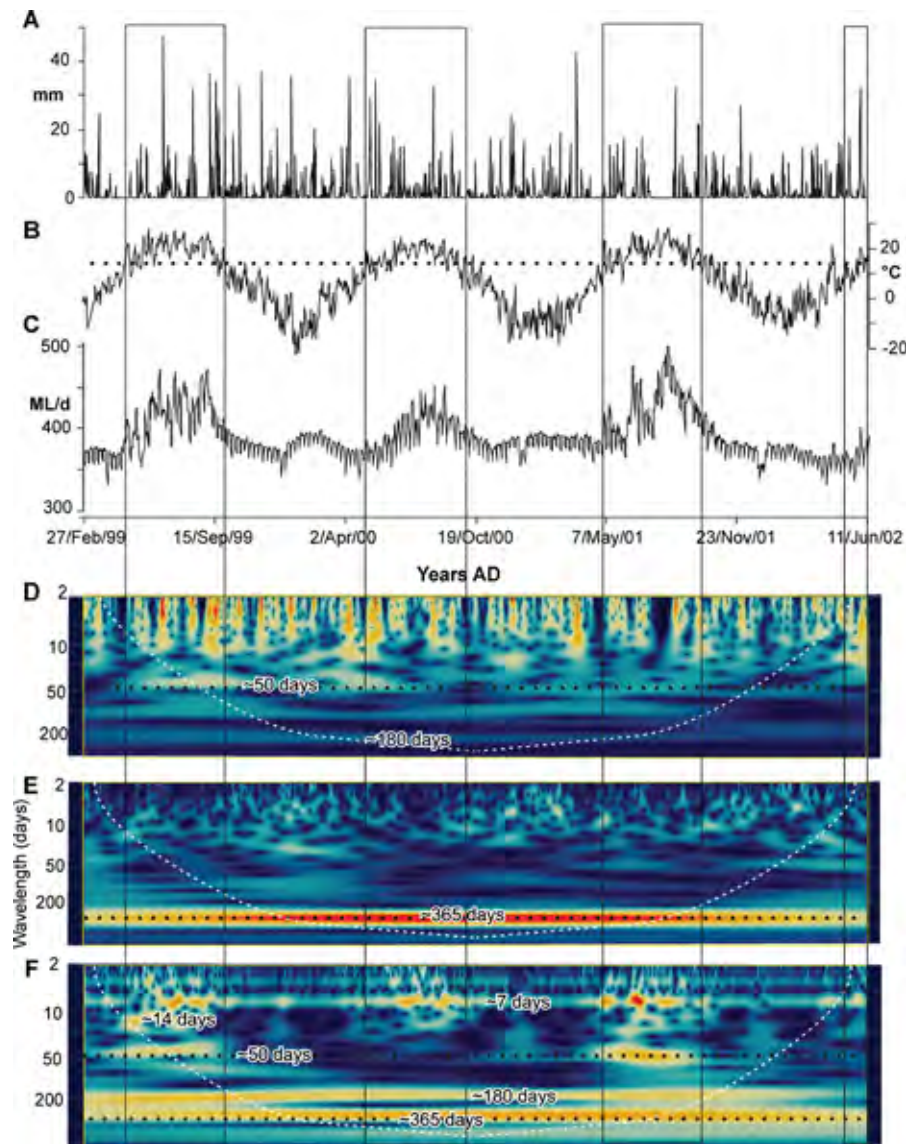
The UWD was most strongly associated with air temperature in Calgary, where UWD increased by approximately 100 ML per day with a 2 °C increase in air temperature. While the correlation between UWD and air temperature was most significant for Montreal, an increase of approximately 100 ML per day in UWD required an increase in air temperature of approximately 10 °C (Table 1). By contrast, no relationship was found between daily records of UWD and total precipitation for any of the cities (Figs. 5D–5F).

Fourier analysis illustrated that the 365-day (annual) wavelength was more than 95 % significant in temperature and UWD for all cities. This indicates that UWD



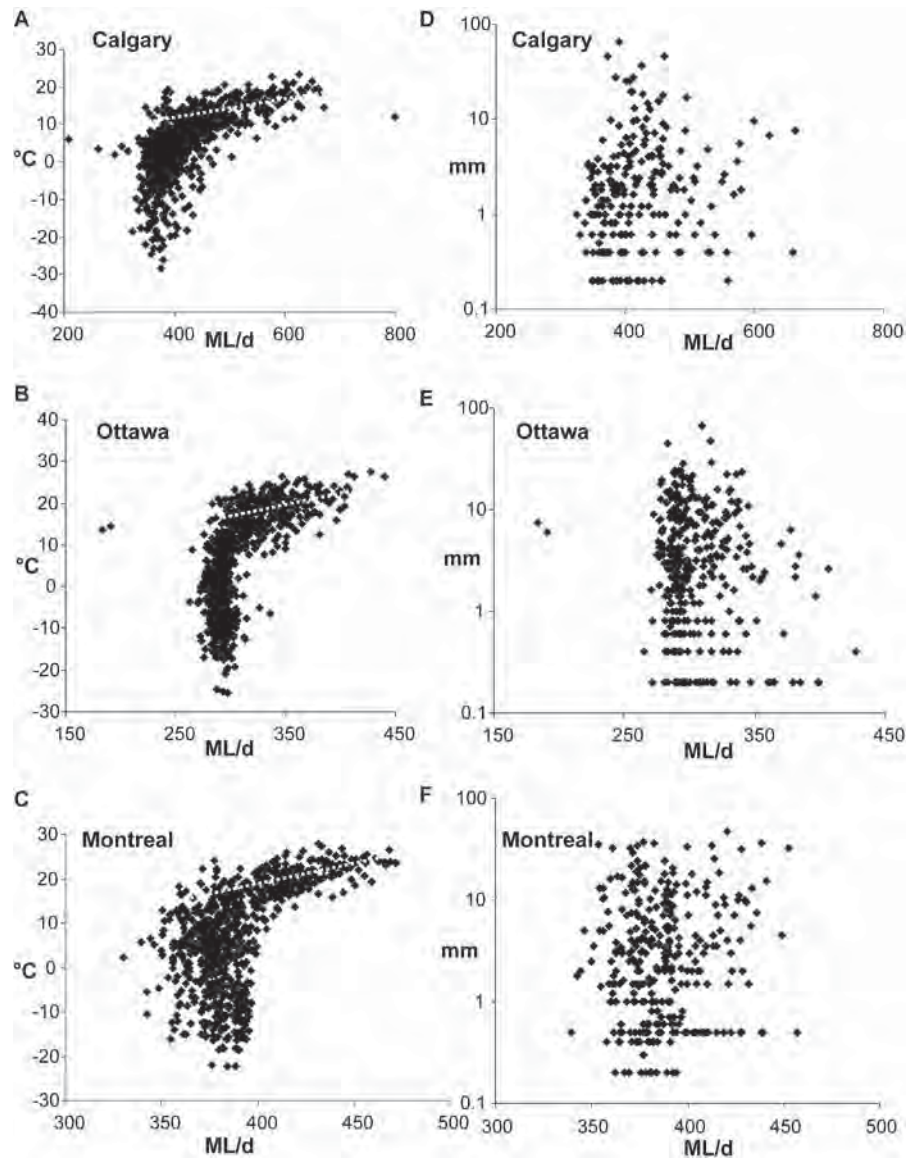
**Fig. 3** Wavelet analysis of daily UWD and meteorological data from Ottawa from 4/2006 to 6/2009: **A** daily total precipitation (mm); **B** daily mean temperature ( $^{\circ}\text{C}$ ) with sensitivity threshold (*dashed line*); **C** urban water demand (ML/d); **D** wavelet scalogram of daily total precipitation record; **E** wavelet scalogram of daily mean temperature record; **F** wavelet scalogram of daily urban water demand. *Curved dashed lines* mark “cone of influence” (see text), *horizontal dotted lines* mark 40-days and 365-days scale, and *boxes* mark high UWD variability intervals. *Orange colors* indicate large wavelet coefficients and *blue colors* weak wavelet coefficients





**Fig. 4** Wavelet analysis of daily UWD and meteorological data from Montreal from 2/1999 to 6/2002: **A** daily total precipitation (mm); **B** daily mean temperature ( $^{\circ}\text{C}$ ) with sensitivity threshold (*dashed line*); **C** urban water demand (ML/d); **D** wavelet scalogram of daily total precipitation record; **E** wavelet scalogram of daily mean temperature record; **F** wavelet scalogram of daily urban water demand. *Curved dashed lines* mark “cone of influence” (see text), *horizontal dotted lines* mark 50-days and 365-days scale, and *boxes* mark high UWD variability intervals. *Orange colors* indicate large wavelet coefficients and *blue colors* weak wavelet coefficients

is strongly dependent on temperature in the three cities studied in Canada (Fig. 6). Fourier analysis also detected that the 7-day cycle related to the working week was significant for all locations, not only in Montreal as indicated by WT. The 3.5 and



**Fig. 5** Scatter plots of daily UWD versus meteorological records: **A–C**—UWD versus mean temperature, *dashed lines* mark regression line above temperature threshold; **D–F**—UWD versus total precipitation

2.3-day cycles may be harmonics (frequency-multiples) of the 7-day cycles, because the 7-day cycle is approximately step-wise (5 days work/2 days no work), and is thus not Gaussian distributed as assumed by Fourier analysis.

Cross-spectral analysis between UWD and meteorological records for the 730-day interval show potential relationships at time offsets, representing a delayed response in the Fourier domain (Fig. 7). As already evident from WT and the records

**Table 1** Daily UWD–mean daily temperature relationship

	Location		
	Calgary	Ottawa	Montreal
Total number of samples	730	730	730
Temperature threshold (°C)	10	12	12
Number of samples above threshold	255	299	291
Regression equation $T = f(\text{UWD})$	$0.021x + 3.95$	$0.06x - 1.71$	$0.101x - 22.56$
Coefficient of determination $R^2$	0.28	0.26	0.49
PEARSON correlation coefficient $r$	0.53	0.51	0.7
Significance level $p$ for $r$	< 0.001	< 0.001	< 0.001

**Table 2** Mean daily temperature–daily UWD and precipitation variance relationships

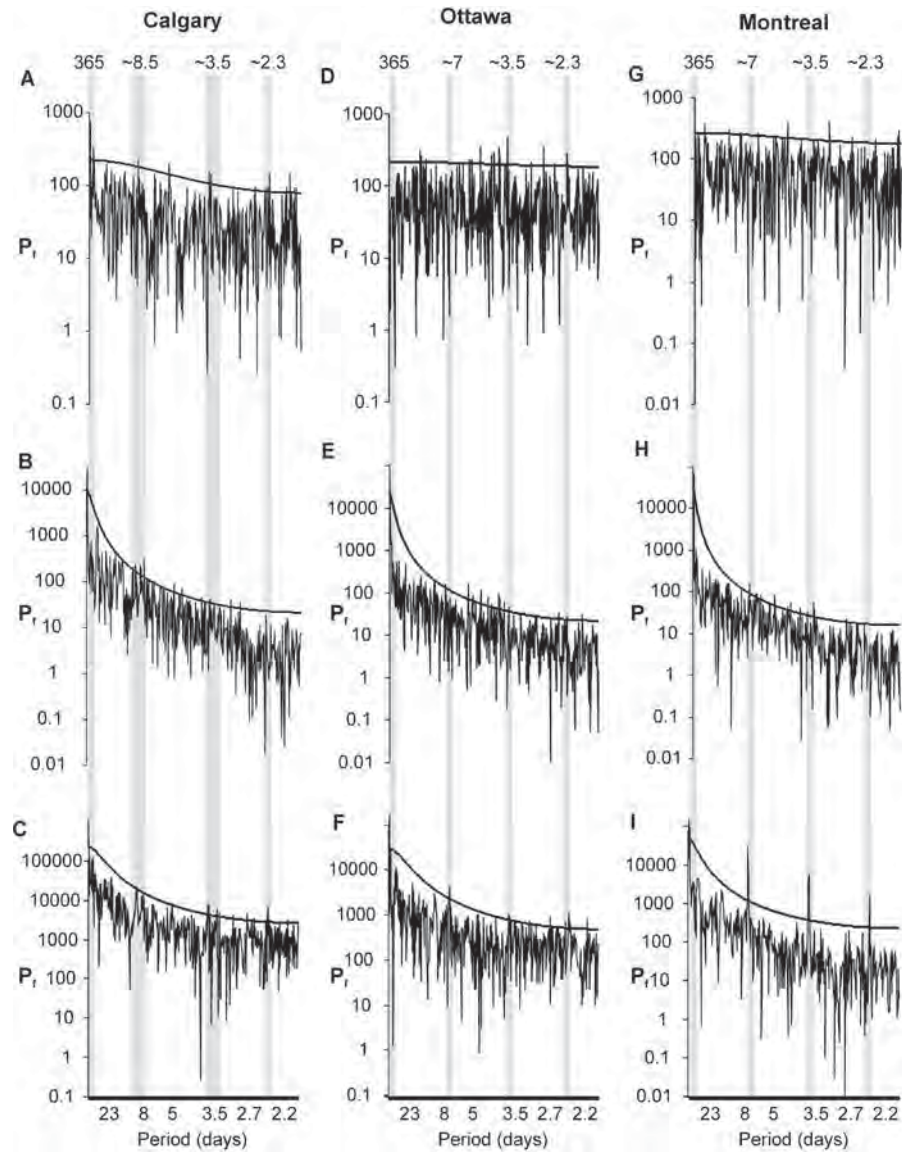
	$T$ range (°C)	UWD variance (ML/d) <sup>2</sup>	Precipitation variance (mm/d) <sup>2</sup>	% of total days
Ottawa	–24 to –6	104	16	18
	–6 to 12	246	31	39
	12 to 30	1043	45	43
Calgary	–26 to –8	496	2	8
	–8 to 10	1467	20	53
	10 to 28	5568	27	39
Montreal	–24 to –6	242	17	16
	–6 to 12	289	45	43
	12 to 30	882	54	41

themselves, there was a significant coherency between UWD and the mean temperature in the annual cycle at  $-0.3$  to  $0.16$  radians, thus indicating only a minor temporal offset (Figs. 7A–7C). Similarly, a strong coherency at  $-0.26$  radian offset also occurred between UWD and precipitation in Calgary, which highlights the correlation between the seasonal patterns of the two parameters.

There were several other significant coherencies between UWD and meteorological records in the 2.2-day to 52-day wavelength spectrum with varying phase offsets (Fig. 7). Most notably, a 7–8.5-day (weekly) cycle occurs between UWD and temperature in all locations, albeit with very different phase offsets that range from positively correlated ( $\sim 0.1$  radians for Montreal) to inverse correlated ( $\sim 2.9$  radians for Ottawa).

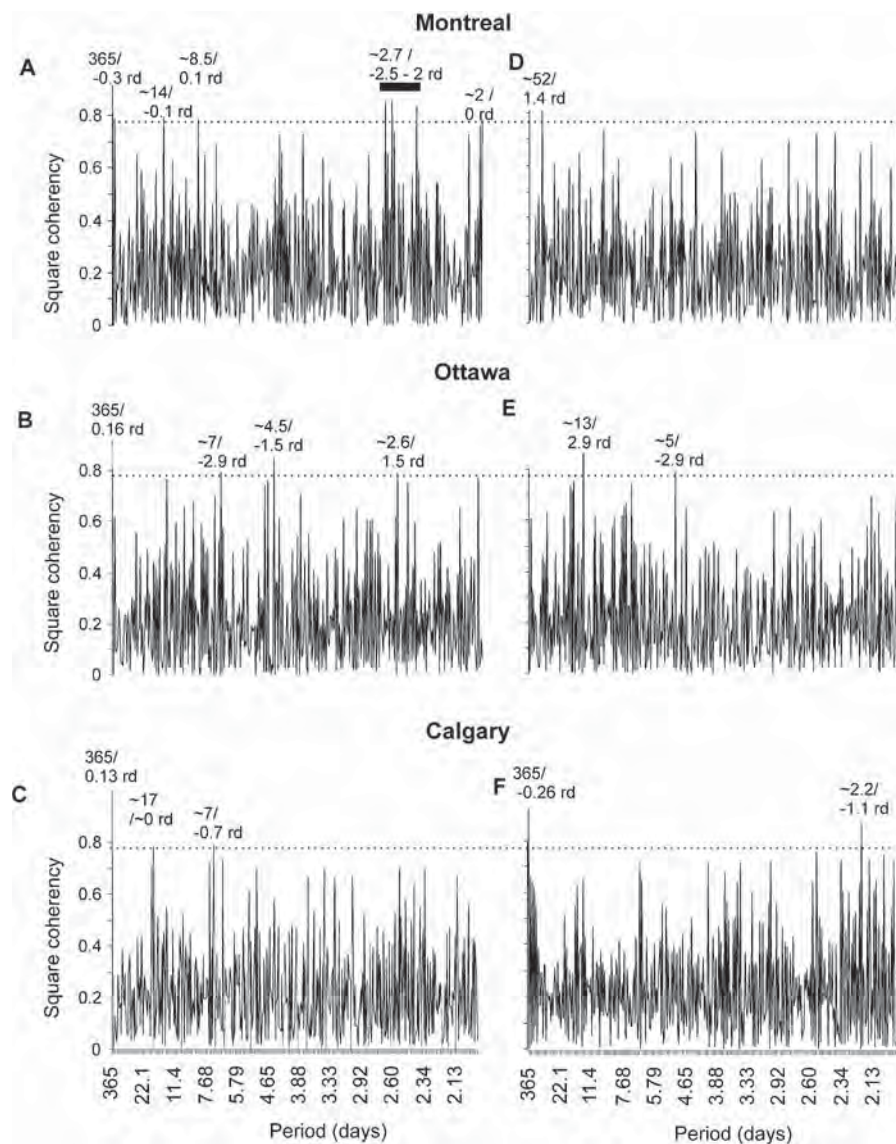
## 6 Discussion

This study applied CWT for the first time to determine changes in the temporal pattern of UWD and its potential meteorological drivers, specifically daily mean air temperature and precipitation, at both frequency and time-scale in three major Canadian



**Fig. 6** Fourier analysis of daily UWD and meteorological data: **A** power spectra of daily total precipitation record from Calgary; **B** power spectra of daily mean temperature record from Calgary; **C** power spectra of daily urban water demand record from Calgary; **D** power spectra of daily total precipitation record from Ottawa; **E** power spectra of daily mean temperature record from Ottawa; **F** power spectra of daily urban water demand record from Ottawa; **G** power spectra of daily total precipitation record from Montreal; **H** power spectra of daily mean temperature record from Montreal; **I** power spectra of daily urban water demand record from Montreal. *Bold lines* mark 95 %  $\chi^2$ -significance of spectral value, *vertical gray bars* highlight major wavelengths (periods)





**Fig. 7** Cross-spectral analysis of daily UWD versus meteorological data: **A–C**—UWD versus mean temperature; **D–F**—UWD versus total precipitation. *Dashed lines* mark squared coherencies  $> 0.78$ , squared coherencies  $> 0.78$  are marked by *period* and phase offset in radians (rd)

cities. This analysis was aided by Fourier and cross-spectral analysis, which show that the annual (365 day) cycle provides the most consistent and significant relationship between UWD and meteorological drivers in southern Canadian cities.

The methods used detected that during the summer UWD is only sensitive to air temperature when mean daily temperatures are higher than 10–12 °C. Above this threshold, the increase in UWD for a 1 °C increase in temperature ranges from 10 ML

per day for Montreal to 50 ML per day for Calgary. In Calgary, where precipitation is lower and more seasonally distributed, the temperature relationship is amplified by the inverse UWD-precipitation relationship that occurs during the summer months. Such temperature threshold dependent UWD variability has been determined by other data analysis methods previously (Gato et al. 2007). However, the wavelet transform simultaneously provides an insight in the (periodic-cyclic) UWD pattern related to the temperature thresholds. These periodic patterns may be used for threshold-and/or seasonal dependent forecasting of UWD.

WT and Fourier analysis also detected 7-day UWD cycles, particularly in the more industrialized and larger city of Montreal (Fig. 4), which are related to fluctuations during the working week with 5-days work and 2-days off (Ruth et al. 2007). In the absence of a known natural forcing mechanism of a 7–8 day temperature cyclicality, the 7–8.5 day cycle between UWD and temperature could be related to an urban heating effect associated with the working week. For example, manufacturing industries with high energy requirements and heat outputs may increase the surrounding air temperature, and may lead to greater UWD, particularly in summer when UWD becomes very sensitive to air temperature. However, each city experiences different phase offsets between the 7-day UWD cycle and the 7-day temperature cycle. Therefore, a potential link between the 7-day UWD cycle related to the working week and a 7-day temperature cycle, such as a working week-related urban heating effect, appears to be elusive at this point. More detailed studies on the UWD heat island interaction may provide more evidence for the existence of such a link. For example, especially in the large manufacturing centers, the UWD and heating variability inside the city is high and can result in micro-climates where UWD interacts and responds in a very localized manner to meteorological changes. Studies could be set up to investigate the significance of such micro-climate interactions. Moreover, water resources managers could also benefit from the application of the wavelet analysis method to determine patterns and relationships on diurnal fluctuations that allow for short-time water supply and control adjustments.

## 7 Conclusion

Forecasting urban water demand is crucial in managing water demand and supply, particularly given the changes that will be incurred by climate change and population growth. Understanding associations between UWD and meteorological factors such as precipitation and air temperature can enable both better forecasting and a deeper understanding of the natural factors that drive urban water demand.

Wavelet analysis was investigated as a potential method for detecting patterns in UWD both in terms of the wavelength of cycles and the time of occurrence of these cycles, which is particularly useful as UWD varies according to season. The study demonstrated that CWT was able to detect cycles in UWD and meteorological factors, and their temporal range in a continuous scale-time resolution space. In contrast, DWT, which has been previously used for UWD forecasting (Adamowski et al. 2012), would not have been able to precisely detect most of the cycles illustrated in this study. For example, at a 1-day data interval record used for WT, DWT would

extract wavelet coefficients at a dyadic resolution such as for  $2^8$  days = 256 days and  $2^9$  days = 512 days, but not for the 365 days = 1 year cycle as was the case in this study.

When the CWT method was tested using data on UWD, air temperature and precipitation for Calgary, Montreal, and Ottawa, it was found that the annual (365 day) cycle dominated the air temperature and UWD patterns for all three cities studied. Fourier analysis was used to support the WT results, and detected a 7-day cycle in UWD in all three cities, which CWT had only detected in Montreal. This demonstrates the usefulness of using multiple approaches in studies such as this one.

A linear correlation between air temperature and UWD above a threshold of 10–12 °C was also identified. Given the strength of this correlation, applying seasonally dependent linear relationships between the UWD and temperature is suggested as perhaps more appropriate and potentially successful for forecasting than applying a continuous nonlinear function to explain all daily fluctuation throughout the entire record. The UWD–precipitation/temperature relationships in this study are derived from daily, and relatively short records, and do not reflect the long-term pattern. On multi-annual to centennial scale features such as population and economic growth, global warming, and fluctuations in oceanic-atmospheric circulation (House-Peters and Chang 2011; Ruth et al. 2007) superimpose the relationships found in this study that are based on daily to annual patterns.

Overall, this study found that WT is effective in detecting cycles in UWD and meteorological parameters, although Fourier analysis should be used in support of the results due to its more sophisticated detection of significance. Furthermore, the inclusion of projections based on linear correlations between air temperature and UWD under consideration of locally-dependent temperature thresholds could improve forecasting of UWD.

**Acknowledgements** This research was funded by a National Science and Engineering Research Council of Canada (NSERC) Grant and a Fonds de Recherche de Quebec—Nature et Technologies New Researcher Grant held by Jan Adamowski.

## References

- Adamowski J, Karapatakis C (2008) Comparison of multivariate regression and artificial neural networks for peak urban water demand forecasting: the evaluation of different ANN learning algorithms. *J Hydrol Eng* 15(10):694–729. doi:[10.1061/\(ASCE\)HE.1943-5584.0000245](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000245)
- Adamowski J, Chan HF, Prasher SO, Ozga-Zielinski B, Sliusarieva A (2012) Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resour Res* 48:W01528. doi:[10.1029/2010WR009945](https://doi.org/10.1029/2010WR009945)
- Adamowski K, Prokoph A, Adamowski J (2009) Development of a new method of wavelet aided trend detection and estimation. *Hydrol Process* 23:2686–2696. doi:[10.1002/hyp.7260](https://doi.org/10.1002/hyp.7260)
- Agthe DE, Billings RB (1980) Dynamic models of residential water demand. *Water Resour Res* 16(3):476–480. doi:[10.1029/WR016i003p00476](https://doi.org/10.1029/WR016i003p00476)
- Anderson R, Miller T, Washburn M (1980) Water savings from lawn watering restrictions during a drought year in Fort Collins, Colorado. *J Amer Water Resour Ass* 16(4):642–645. doi:[10.1111/j.1752-1688.1980.tb02443.x](https://doi.org/10.1111/j.1752-1688.1980.tb02443.x)
- Arbués F, Barberán R, Villanúa I (2004) Price impact on urban residential water demand: a dynamic panel data approach. *Water Resour Res* 40:W11402. doi:[10.1029/2004WR003092](https://doi.org/10.1029/2004WR003092)

- Arbués F, Villanúa I, Barberán R (2010) Household size and residential water demand. *Aust J Agric Resour Econ* 54:61–80. doi:10.1111/j.1467-8489.2009.00479.x
- Aubertin L, Aubin A, Pelletier G, Curodeau D, Osseyrane M, Lavalée P (2002) Identifying and prioritizing infrastructure rehabilitation. North American Society for Trenchless Technology, Liverpool
- Bassett K, Tinline R (1970) Cross-Spectral analysis of time series and geographical research. *Area* 2:19–24
- Coulbaly P (2006) Spatial and temporal variability of Canadian seasonal precipitation (1900–2000). *Adv Water Resour* 29:1846–1865. doi:10.1016/j.advwatres.2005.12.013
- City of Calgary (2012) Calgary's water demand. [www.calgary.ca/UEP/Water/Pages/Water-conservation/Calgarys-water-demand.aspx](http://www.calgary.ca/UEP/Water/Pages/Water-conservation/Calgarys-water-demand.aspx). Accessed 25th February 2012
- City of Ottawa (2010) Water efficiency plan annual review 2010. Ottawa, ON
- Davis JC (2002) *Statistics and data analysis in geology*, 3rd edn. Wiley, New York. 637 pp
- Environment Canada (2010) Environmental trends. CESI 1 (3)
- Environment Canada (2012) [http://climate.weatheroffice.gc.ca/climateData/canada\\_e.html](http://climate.weatheroffice.gc.ca/climateData/canada_e.html). Accessed 7th January 2012
- Gato S, Jayasuriya N, Roberts P (2007) Temperature and rainfall thresholds for base use urban water demand modelling. *J Hydrol* 337(3–4):364–376. doi:10.1016/j.jhydrol.2007.02.014
- House-Peters LA, Chang H (2011) Urban water demand modeling: review of concepts, methods, and organizing principles. *Water Resour Res* 47:W05401. doi:10.1029/2010WR009624
- IPCC (2007) Climate change 2007: working group I report: the physical science basis. In: Solomon S, et al (eds) Fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, UK and New York, USA
- Jenkins GM, Watts DG (1968) *Spectral analysis and its applications*. Holden-Day, New York
- Maidment D, Parzen E (1984) Monthly water use and its relationship to climatic variables in Texas. *Water Resour Bull* 19(8):409–418
- Maidment DR, Miaou SP (1986) Daily water use in nine cities. *Water Resour Res* 22(6):845–885. doi:10.1029/WR022i006p00845
- Maidment D, Miaou S, Crawford M (1985) Transfer function models of daily urban water use. *Water Resour Res* 21(4):425–432. doi:10.1029/WR021i004p00425
- Mann MS, Lees JM (1996) Robust estimation of background noise and signal detection in climatic time series. *Clim Change* 33:409–445. doi:10.1007/BF00142586
- Maraun D, Kurths J (2004) Cross wavelet analysis: significance testing and pitfalls. *Nonlinear Process Geophys* 11:505–514. doi:10.5194/npg-11-505-2004
- Maraun D, Kurths J, Holschneider M (2007) Nonstationary Gaussian processes in wavelet domain: synthesis, estimation, and significance testing. *Phys Rev E* 75:016707-1–13. doi:10.1103/PhysRevE.75.016707
- Martínez-Espiñeira R (2002) Residential water demand in the northwest of Spain. *Environ Resour Econ* 21(2):161–187
- Miaou SP (1990) A class of time series urban water demand models with nonlinear climatic effects. *Water Resour Res* 26(2):169–178. doi:10.1029/WR026i002p00169
- Morlet J, Arehs G, Fourgeau I, Giard D (1982) Wave propagation and sampling theory. *Geophysics* 47:203–236. doi:10.1190/1.1441328
- Patterson RT, Prokoph A, Chang A (2004) Late Holocene sedimentary response to solar and cosmic ray activity influenced climate variability in the NE Pacific. *Sediment Geol* 172:67–84. doi:10.1016/j.sedgeo.2004.07.007
- Praskievicz S, Chang H (2009) Identifying the relationships between urban water consumption and weather variables in Seoul, South Korea. *Phys Geogr* 30(4):324–337. doi:10.2747/0272-3646.30.4.324
- Prokoph A, Barthelmes F (1996) Detection of nonstationarities in geological time series: wavelet transform of chaotic and cyclic sequences. *Comput Geosci* 22:1097–1108. doi:10.1016/S0098-3004(96)00054-4
- Rioul O, Vetterli M (1991) Wavelets and signal processing. *IEEE Signal Process Mag* 8(4):14–38. doi:10.1109/79.91217
- Ruth M, Bernier C, Jollands N, Golubiewski N (2007) Adaptation of urban water supply infrastructure to impacts from climate and socioeconomic changes: the case of Hamilton, New Zealand. *Water Res Man* 21:1031–1045. doi:10.1007/s11269-006-9071-x
- Schaeffli B, Maraun D, Holschneider M (2007) What drives high flow events in the Swiss Alps? Recent developments in wavelet spectral analysis and their application to hydrology. *Adv Water Resour* 30:2511–2525. doi:10.1016/j.advwatres.2007.06.004



- Schleich J, Hillenbrand T (2009) Determinants of residential water demand in Germany. *Ecol Econ* 68:1756–1769. doi:[10.1016/j.ecolecon.2008.11.012](https://doi.org/10.1016/j.ecolecon.2008.11.012)
- Schulz M, Mudelsee M (2002) REDFIT: estimating red-noise spectra directly from unevenly spaced paleoclimatic time series. *Comput Geosci* 28:421–426. doi:[10.1016/S0098-3004\(01\)00044-9](https://doi.org/10.1016/S0098-3004(01)00044-9)
- Torrence C, Compo GP (1998) A practical guide to wavelet analysis. *Bull Am Meteorol Soc* 79:61–78. doi:[10.1175/1520-0477\(1998\)079<0061:APGTWA>2.0.CO;2](https://doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2)
- Ware DM, Thomson RE (2000) Interannual to multidecadal timescale climate variations in the Northeast Pacific. *J Climate* 13:3209–3220. doi:[10.1175/1520-0442\(2000\)013<3209:ITMTCV>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<3209:ITMTCV>2.0.CO;2)
- Zhang X, Vincent A, Hogg WD, Niitsoo A (2000) Temperature and precipitation trends in Canada during the 20th century. *Atmos-Ocean* 38:395–429. doi:[10.1080/07055900.2000.9649654](https://doi.org/10.1080/07055900.2000.9649654)
- Zhou S, McMahon T, Walton A, Lewis J (2000) Forecasting daily urban water demand: a case study of Melbourne. *J Hydrol* 236:153–164. doi:[10.1016/S0022-1694\(00\)00287-0](https://doi.org/10.1016/S0022-1694(00)00287-0)

## Assessing the Impacts of Four Land Use Types on the Water Quality of Wetlands in Japan

Azam Haidary · Bahman Jabbarian Amiri ·  
Jan Adamowski · Nicola Fohrer · Kaneyuki Nakane

Received: 5 September 2012 / Accepted: 20 January 2013 /  
Published online: 2 February 2013  
© Springer Science+Business Media Dordrecht 2013

**Abstract** This study examined how changes in the composition of land use can affect wetland water quality. Twenty-four wetlands located in Hiroshima prefecture in the western part of Japan were selected for this purpose. The water quality parameters that were explored include: pH, electrical conductivity, turbidity, dissolved oxygen, total dissolved solid, temperature and different forms of nitrogen. These important indicators of the water quality in the study area were measured from December 2005 to December 2006. The composition of land uses was determined for the catchments of the wetlands. They were then categorized into three classes, including non-disturbed, moderately-disturbed and highly-disturbed wetlands, based on the extent of urban area (as the most disruptive land use type within the catchment of the wetlands). The relationship between land use types and water quality parameters for the wetlands was statistically examined. The findings indicated that there were significant positive relationships between the proportion (%) of urban areas within catchments of the wetlands and EC ( $r=0.67$ ,  $p<0.01$ ), TDS ( $r=0.69$ ,  $p<0.01$ ), TN ( $r=0.92$ ,  $p<0.01$ ), DON ( $r=0.6$ ,  $p<0.01$ ),  $\text{NH}_4^+$  ( $r=0.47$ ,  $p<0.05$ ),  $\text{NO}_2^-$  ( $r=0.50$ ,  $p<0.05$ ), while negative relationships were observed between the proportion (%) of forest area in these wetlands and EC ( $r=-0.62$ ,  $p<0.01$ ), TDS ( $r=-0.68$ ,  $p<0.01$ ), TN ( $r=-0.68$ ,  $p<0.01$ ), DON ( $r=-0.43$ ,  $p<0.05$ ), and  $\text{NH}_4^+$  ( $r=-0.55$ ,  $p<0.01$ ). Analysis of the variance also

---

A. Haidary · K. Nakane  
Division of Environmental Dynamics and Management, Graduate School of Biosphere Science,  
Hiroshima University, 1-7-1 Kagamiyama, Higashi-Hiroshima 739-8521, Japan

B. J. Amiri (✉)  
Department of Environmental Science, Faculty of Natural Resources, University of Tehran, Karaj,  
P.O. Box: 4314, Iran  
e-mail: j.amiri@yahoo.com

J. Adamowski  
Department of Bioresource Engineering, Faculty of Agricultural and Environmental Sciences, McGill  
University, Montreal, Quebec, Canada

N. Fohrer  
Department of Hydrology and Water Resources Management, Ecology Centre, Institute of Nature  
Protection and Water Resources Management, Christian Albrecht Universität zu Kiel, Olshausenstrasse  
75, Geb. I, 24118 Kiel, Germany

revealed significant differences within the wetland groups in terms of the annual mean of electrical conductivity, total dissolved solids, total nitrogen, nitrite, dissolved inorganic nitrogen and dissolved organic nitrogen in the study area. Moreover, the study also indicated that the forest area plays a significant role in withholding nutrient loads from the wetlands, and hence, it can act as a sink for surface/subsurface nutrient inputs flowing into such water bodies from the watersheds.

**Keywords** Wetland · Land use · Water quality · Catchment

## 1 Introduction

Aquatic ecosystems in general, and wetlands in particular, have been used by humans over the centuries to the extent that not that many have remained today in their natural condition, as a result of pollution loads, among other reasons (Ngoye and Machiwa 2004). The water quality of water resources is generally linked with land use within a catchment since a catchment can influence the quality and quantity of runoff during and after a rainfall event (Richards and Host 1994). Hence, a lot of water pollution problems are caused by changes in the composition of land use within a catchment as human activities increase (Gikas et al. 2006; Amiri and Nakane 2009; Boskidis et al. 2011). Among different aquatic ecosystems, wetlands not only play a key role in water quality improvement of other water bodies such as rivers at the catchment-scale, but they are also themselves influenced by changes in the composition of land uses due to human activities within their catchments (Plameri and Treppel 2002; Papastergiadou et al. 2008). Removal of nutrient loads from running waters is an important role of wetlands in watersheds (Jones et al. 2001; Jordan et al. 2003; Akrotos et al. 2006; Plameri and Treppel 2002); these types of water bodies are often called the “kidney” of a watershed (Brooks et al. 2003). The nutrient removal ability of wetlands is based on trapping sediment, removing nutrients, storing and releasing inorganic nutrients and transforming them into organic forms (DeBusk 1999). Due to their significant role in the improvement of the environmental quality of catchments, significant attention has been given to the conservation and restoration of natural wetlands, as well as the construction of artificial ones in catchments (Tsihrintzis et al. 1995; Lee et al. 2005).

Due to the degradation in wetland ecosystems resulting from changes in the composition of land use of their catchments (Papastergiadou et al. 2008), significant research has been conducted (e.g., Schueler 1994; Arnold and Gibbons 1996; Tsihrintzis and Hamid 1997; May et al. 1997; Brabec et al. 2002; Clapcott et al. 2011) to explore the relationship between urban area (impervious cover) and non-point source pollution, in addition to documenting the adverse impacts on water quality. Although pollution is not generated by impervious covers, many of the physical and biological impacts affecting the quality of water resources, such as streams and wetlands, originate due to the impervious cover-induced hydrological changes (May et al. 1997). The imperviousness of the urban areas increases their hydrological activity, and even small rainfalls are capable of washing the accumulated pollutants into water bodies (Basnyat et al. 1999). Regarding the assessment of water quality conditions of streams, several authors (e.g., Schueler 1994; Arnold and Gibbons 1996; May et al. 1997; Helms et al. 2009; Gregory and Calhoun 2007) have suggested a linkage between the impervious cover and in-stream water quality degradation. Out of the afore-cited works, Schueler (1994) has described the impacts of an increase in the impervious cover on in-stream water quality as occurring in three ascending levels namely, less than 10 % (non-disturbed streams), 10~25 % (semi-disturbed streams) and more than 25 % (disturbed

streams) within the catchment of a river. Moreover, Amiri et al. (2012) has determined a threshold for two disturbing land uses, namely, agricultural area (60 %) and grassland area (10 %) within the catchment of the rivers in relation to total phosphorus, so that when exceeding those thresholds, the in-stream total phosphorus concentration increases significantly. May et al. (1997) suggest a 10 % impervious cover threshold that can be applied to wetland communities. Moreover, an inverse correlation between wetland habitat quality and increasing impervious cover has also been reported by Clapcott et al. (2011), which is consistent with the 10 % impervious cover threshold as suggested by May et al. (1997). Although there have been some studies in which the relationship between a change in the composition of land uses and variations in the water quality parameters of wetlands has been examined (e.g.; Tsihrintzis et al. 1996, 1997; Crosbie and Chow-Fraser 1999; Daley and McDowell 2002; Houlihan and Findley 2004; Haidary and Nakane 2008), few studies have applied an ordination approach to specify how changes in land use types in the catchment of the wetlands affects their water quality. Accordingly, the objective of this study is firstly to assess whether different types of human activities expressed through four land uses (i.e., urban, forest, agricultural and grassland) within the catchment of the wetland can affect water quality parameters, and secondly whether a threshold of water quality degradation might be determined considering changes in the percentage of urban area as a disturbing land use type in the catchment of the wetlands of interest.

## 2 Materials and Methods

### 2.1 Study Area

The study area of this research project is located in Higashi-Hiroshima in Japan, which is placed within 132° 36' 23"~132° 51' 19" E and 34° 15' 19"~34° 34' 58" N, with an area of 635 km<sup>2</sup> (Fig. 1). Annual rainfall is, on average, 160 mm/month, with a maximum monthly value of 304 mm and a minimum monthly value of 19 mm, which were recorded in July and October in the study area, respectively. Although mean annual temperature is 14.1 °C, the monthly mean varies from 2.3 °C (in January) to 26 °C (in August) (Japan Meteorological Agency). Based on spatial analysis, which was conducted using geological formation, soil and land-use maps of the wetlands, it was found that granite and alluvial sand are the main geological formations in the study area. Dominant and sub-dominant soil types are residual regosols and brown forest soil, respectively. Figure 1 depicts the geographical distribution of the 24 wetlands explored in this study. It is noteworthy that all the wetlands have similar geographic features, except the physical features of the watersheds. Twenty-four out of 1,100 wetlands, as reported by Shimoda (1993), were chosen as our study sites using a topographical map (1:50,000) (Japan Geographical Survey Institute). Table 1 indicates geometric features, the proportion (%) of land use types and that of soil types and geological formations for the catchment of each wetland.

### 2.2 Materials and Methods

Hydrological modeling is an efficient tool to explore the relationship between entities in a given catchment (e.g.: Boskidis et al. 2011); statistical approaches are also helpful when one has an appropriate amount of data and no adequate knowledge





**Fig. 1** Geographical position of the study area

regarding the entities of interest (Grant et al. 1997). In this study, a statistical approach was applied. According to this approach, environmental impacts of human activities are statistically assessed by partitioning the ecosystem of interest into homogeneous environmental impact units considering the extent of the human disturbances, which may occur across the ecosystem of interest (Liddle 1975). These homogenous environmental impact units are then called the disturbed, semi-disturbed, and non-disturbed environmental impact units, respectively. These ecological impact units can also statistically be considered as treatments, and environmental impacts of the human disturbances can statistically be analyzed within and between these spatial treatments.

A variety of water quality parameters in the study area wetlands, namely pH, electrical conductivity (EC mS/m), turbidity, dissolved oxygen (DO mg/L), temperature (C), and total dissolved solids (TDS mg/L) were measured by a portable water quality monitoring device (HORIBA, Model U-21 XD) in the inflows of each wetland throughout the course of four seasons in 2006 (Table 2). In flowing surface were simultaneously sampled and immediately placed in a cooler in order to transport to the laboratory for determining different forms of nitrogen including nitrate ( $\text{NO}_3^-$ ), nitrite ( $\text{NO}_2^-$ ) (Ion Chromatography Method) and ammonium ( $\text{NH}_4^+$ ) (APHA 1995), dissolved inorganic nitrogen (DIN), dissolved organic nitrogen (DON) (Semi-Micro Kjeldahl Method), and total nitrogen (TN) (4500-Norg C).

All water quality parameter data was tested for normality using the Shapiro–Wilk test with a p-value of less than 0.05 (Table 3). The Spearman rank correlation test was then applied to determine if any of the water quality variables were associated with changes in percentage of land use, those of soil types and geological formations along with geometric features namely, catchment area (hac), average catchment slope (%), average main channel slope (%), drainage density (m/hac) and Strahler order within catchment of the wetlands (Table 4).

**Table 1** Geometric features, soil type and geological formations of the catchment of the wetlands in study area

Wetland no.	Geometrics of the catchments				Proportion (%) of soil type					Proportion (%) of geological formation			
	Catchment area (hac)	Average catchment slope (%)	Strahler order	Drainage density (m/hac)	Average main channel slope (%)	Gray lowland soil	Residual regosol	Brown forest soil	Brown forest soil (dry)	Regosol	Diluvial sand	Granite	Rhyolite
1	39.43	12.14	2	58.39	12.14	0.00	0.00	0.00	0.00	0.00	82.63	0.00	0.00
2	9.43	18.12	2	65.18	18.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	2.53	10.21	1	17.01	10.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	14.74	4.48	2	47.76	4.48	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	14.38	2.75	2	51.82	2.75	5.00	0.00	7.34	91.00	0.00	82.89	17.11	0.00
6	27.45	17.33	2	42.51	17.33	0.00	0.00	100	0.00	0.00	0.00	0.00	0.00
7	17.07	7.54	2	51.45	7.54	0.00	5.40	94.60	15.02	0.00	0.00	84.73	15.27
8	330.87	14.76	4	49.56	14.76	0.00	1.00	0.00	0.00	0.00	16.72	83.28	0.00
9	4.5	1.92	1	20.32	1.92	0.00	0.00	0.00	0.00	1.32	0.00	9.72	90.28
10	42.98	11.78	2	46.58	11.78	0.00	0.00	0.00	0.00	17.72	0.00	0.06	99.94
11	4.14	15.75	1	26.69	15.75	0.00	0.00	0.00	0.00	0.00	0.26	97.92	0.00
12	2.55	20.83	1	44.46	20.83	0.00	0.00	0.00	0.00	100	57.63	6.46	35.92
13	4.52	7.89	1	26.32	7.89	0.00	0.00	89.17	0.00	0.00	100	94.11	5.63
14	150.78	6.35	3	44.54	6.35	50.43	0.00	26.15	23.42	0.00	19.63	80.37	0.00
15	3.09	4.19	1	35.05	4.19	0.00	79.35	0.00	7.92	0.00	0.00	61.91	9.41
16	34.97	3.18	2	53.86	3.18	0.00	1.66	0.00	48.32	0.00	45.83	54.17	0.00
17	44.08	1.66	3	47.64	1.66	27.28	72.72	0.00	0.00	0.00	1.39	98.61	0.00
18	2.78	10.39	1	39.64	10.39	0.00	100	0.00	24.63	0.00	72.83	27.17	0.00
19	21.30	9.41	2	30.94	7.53	0.00	8.62	91.38	0.00	0.00	0.00	27.08	27.92
20	14.53	11.74	2	28.87	11.74	1.02	18.38	80.60	0.00	0.00	25.26	74.74	0.00
21	9.40	5.28	1	41.06	5.28	0.00	72.77	27.73	0.00	0.00	28.68	100.00	0.00
22	5.08	3.08	1	29.59	3.08	18.83	81.17	0.00	0.00	0.00	47.84	52.16	0.00
23	26.23	10.28	2	46.29	10.28	100	0.00	0.00	0.00	0.00	2.08	17.37	0.00
24	45.35	8.62	3	46.04	8.62	100	0.00	0.00	0.00	0.00	100	0.00	0.00

**Table 2** Annual mean of water quality parameters of the wetlands

Wetlands	Water quality parameters											
	pH	Ec (mS/m)	Turbidity (mg/L)	DO (mg/L)	TDS (mg/L)	T (C)	TN (mg/L)	NO <sub>3</sub> <sup>-</sup> (mg/L)	DON (mg/L)	NH <sub>4</sub> <sup>+</sup> (mg/L)	NO <sub>2</sub> <sup>-</sup> (mg/L)	DIN (mg/L)
1	6.27	12.33	15.00	6.60	0.08	18.44	0.75	0.03	0.68	0.007	0.006	0.06
2	6.50	4.00	12.75	6.95	0.03	18.20	0.55	0.03	0.52	0.003	0.005	0.03
3	6.20	11.75	9.75	5.98	0.08	16.63	0.93	0.02	0.89	0.006	0.003	0.05
4	6.28	14.00	23.00	7.43	0.10	18.60	0.60	0.06	0.51	0.016	0.007	0.08
5	6.63	15.75	9.75	5.65	0.14	17.95	1.06	0.15	0.66	0.232	0.021	0.40
6	6.48	7.00	7.25	5.10	0.05	19.15	0.60	0.02	0.57	0.004	0.004	0.03
7	6.85	5.00	21.5	9.35	0.03	17.40	0.49	0.02	0.47	0.000	0.004	0.02
8	6.75	6.75	7.00	4.45	0.05	18.73	0.47	0.03	0.43	0.005	0.002	0.04
9	6.53	4.50	3.00	7.38	0.03	18.80	0.50	0.04	0.46	0.001	0.002	0.04
10	6.40	10.50	8.25	7.65	0.05	17.93	0.49	0.21	0.27	0.007	0.007	0.22
11	6.58	9.50	20.50	6.93	0.06	18.38	0.64	0.15	0.40	0.082	0.005	0.24
12	6.88	8.50	14.25	7.90	0.07	16.43	0.42	0.08	0.34	0.005	0.003	0.08
13	6.35	7.75	11.75	4.73	0.05	18.35	0.63	0.05	0.56	0.018	0.003	0.07
14	6.40	9.00	18.00	6.78	0.06	19.90	0.89	0.03	0.77	0.009	0.015	–
15	6.57	8.00	16.00	6.53	0.05	19.48	0.66	0.03	0.41	0.004	0.004	0.04
16	6.64	16.50	22.25	7.35	0.11	19.28	1.28	0.25	0.90	0.099	0.019	0.37
17	6.30	13.00	31.75	6.50	0.09	17.80	0.87	0.08	0.70	0.06	0.027	0.17
18	6.19	8.00	8.75	6.10	0.05	18.80	0.51	0.03	0.48	0.007	0.000	0.02
19	6.44	6.50	14.50	6.13	0.04	19.65	0.42	0.04	0.37	0.006	0.002	0.05
20	6.44	7.00	16.00	6.80	0.05	18.73	0.46	0.05	0.40	0.007	0.002	0.06
21	6.38	8.25	23.25	6.95	0.06	19.05	0.78	0.02	0.75	0.006	0.001	0.03
22	6.07	4.75	12.20	10.22	0.03	13.13	0.42	0.07	0.32	0.025	0.005	0.10
23	6.63	5.33	52.30	7.84	0.04	16.50	0.63	0.06	0.48	0.069	0.023	0.15
24	6.51	15.25	18.75	5.34	0.10	16.26	2.26	0.65	0.36	1.164	0.086	1.90

Considering the landscape degradation within the study wetlands, they were classified into three groups in order to assess the impact of landscape degradation on the water quality of the wetlands (Fig. 2). The classification of the wetlands was conducted considering the percentage of urban areas within the catchment of each wetland. The Kruskal–Wallis test was then applied to specify whether there were significant differences among the three wetland groups in terms of the annual mean of water quality parameters (Table 5).

The watershed boundary of the wetlands was defined by applying 30 m DEM by the Spatial Analyst extension in ArcMap 9.3. The land use map (Amiri and Nakane 2006) was then overlaid on the catchment boundary of the wetlands to calculate the real extent of each land use type within the catchments. The result of this calculation was then divided by the area of the catchments to determine the percentage of the catchment covered by each type of land use by Geographical Information System (ArcMap 9.3). The same procedure was also applied to calculate the proportion (%) of soil type and geological formations for each of the wetlands.

**Table 3** Result of normality test for water quality parameters

Parameter	Wilk–Shapiro test		
	Statistic	Significance	
Water quality	pH	0.99	0.98 <sup>a</sup>
	EC (mS/m)	0.95	0.09 <sup>a</sup>
	Turbidity (mg/L)	0.83	0.00
	DO (mg/L)	0.96	0.47 <sup>a</sup>
	TDS (mg/L)	0.89	0.01
	T (C)	0.83	0.00
	TN (mg/L)	0.67	0.00
	NO <sub>3</sub> <sup>-</sup> (mg/L)	0.56	0.00
	DON (mg/L)	0.93	0.09 <sup>a</sup>
	NH <sub>4</sub> <sup>+</sup> (mg/L)	0.34	0.00
	NO <sub>2</sub> <sup>-</sup> (mg/L)	0.52	0.00
	DIN (mg/L)	0.42	0.00
Geometrics	Elevation (m)	0.94	0.12 <sup>a</sup>
	Catchment area (hec)	0.48	0.00
	Strahler order	0.82	0.00
	Drainage density (m/hect)	0.97	0.59 <sup>a</sup>
	Main channel slope	0.94	0.16 <sup>a</sup>
Land covers	Catchment slope	0.94	0.57 <sup>a</sup>
	Urban	0.72	0.00
	Forest	0.88	0.10 <sup>a</sup>
	Agriculture	0.89	0.10 <sup>a</sup>
Soils	Grassland	0.77	0.00
	Grey lowland soil	0.49	0.00
	Residual regosol	0.59	0.00
	Brown forest soil	0.61	0.00
Geology	Regosol	0.49	0.00
	Diluvial sand	0.26	0.00
	Granite	0.78	0.00
	Rhyolite	0.84	0.00

<sup>a</sup>Correlation is significant at the 0.01 level (2-tailed)

### 3 Results and Discussion

The relationship between the annual mean value of the water quality parameters of the wetlands and the proportion of land use types within the watersheds of the wetlands was examined using the Spearman correlation coefficient test ( $p < 0.05$ ). The findings revealed that there were significant positive associations between the proportion (%) of urban areas within the watersheds of the wetlands and the EC value ( $r = 0.67$ ,  $p < 0.01$ ), TDS ( $r = 0.69$ ,  $p < 0.01$ ), TN ( $r = 0.92$ ,  $p < 0.01$ ), DON ( $r = 0.6$ ,  $p < 0.01$ ), NH<sub>4</sub><sup>+</sup> ( $r = 0.47$ ,  $p < 0.05$ ), NO<sub>2</sub><sup>-</sup> ( $r = 0.50$ ,  $p < 0.05$ ), while negative associations were observed between DO ( $r = -0.42$ ,  $p < 0.05$ ), and % of urban area in the watershed of the wetlands. This could be the result of an increase in nutrient concentrations. Accordingly, if the proportion (%) of the urban area increases within the catchment of the wetlands, the annual mean of the water quality parameters,



**Table 4** Spearman correlation coefficient test between water quality parameters and proportion (%) of land use, catchment geometrics, and soil and geological features in the catchment of the wetlands

Water quality parameter	Land uses			Catchment geometrics				Soil and geology			
	Urban	Forest	Agriculture	Grassland	Catchment area	Strahler order	Drainage density	Grey lowland soil	Regosol	Diluvial sand	Rhyolite
Ec (mS/m)	0.67 <sup>a</sup>	-0.62 <sup>a</sup>	0.40 <sup>b</sup>	0.06	-	-	-	-	-	-	-
DO (mg/L)	-0.42 <sup>b</sup>	0.34	-0.17	-0.06	-	-	-	-	0.42	-	-
TDS (mg/L)	0.69 <sup>a</sup>	-0.68 <sup>a</sup>	0.45 <sup>b</sup>	0.15	-	-	-	-	-	0.41 <sup>b</sup>	-
TN (mg/L)	0.92 <sup>a</sup>	-0.68 <sup>a</sup>	0.44 <sup>b</sup>	-0.05	-	-	-	-	-	-	0.46 <sup>b</sup>
DON (mg/L)	0.65 <sup>a</sup>	-0.43 <sup>b</sup>	0.20	0.04	-	-	-	-	-0.44 <sup>b</sup>	-	-0.49 <sup>b</sup>
NH <sub>4</sub> <sup>+</sup> (mg/L)	0.47 <sup>b</sup>	-0.55 <sup>a</sup>	0.56 <sup>a</sup>	0.03	-	-	-	0.60 <sup>a</sup>	-	0.58 <sup>a</sup>	-0.47 <sup>b</sup>
NO <sub>2</sub> <sup>-</sup> (mg/L)	0.50 <sup>b</sup>	-0.39	0.31	-0.09	0.54 <sup>a</sup>	0.52 <sup>a</sup>	0.53 <sup>a</sup>	0.59 <sup>a</sup>	-	-	-

<sup>a</sup> Correlation is significant at the 0.01 level (2-tailed)<sup>b</sup> Correlation is significant at the 0.05 level (2-tailed)

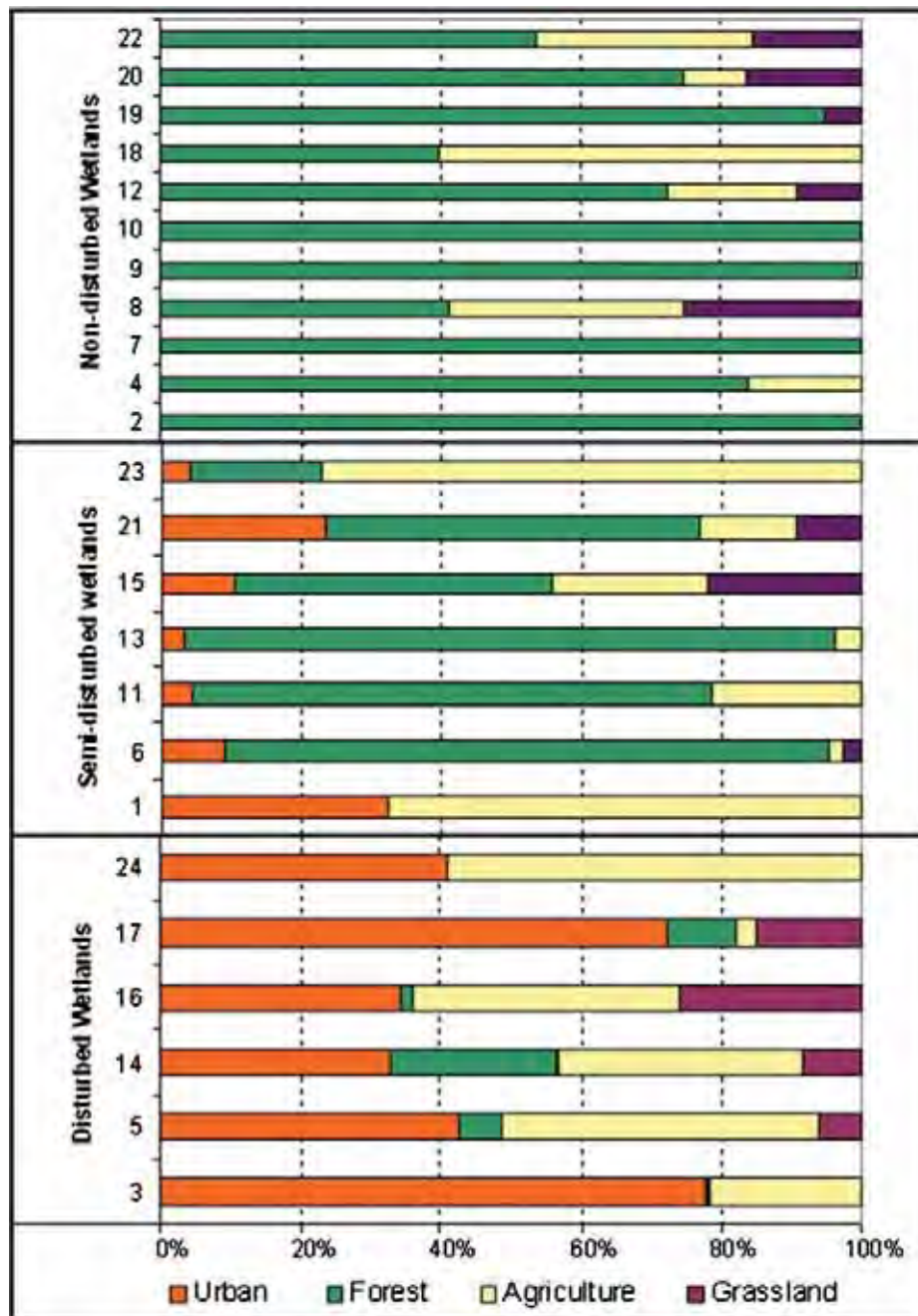


Fig. 2 Proportion (%) of land use in the catchment of the wetlands along with their classification into three classes: non-disturbed, semi-disturbed and disturbed wetlands

except DO, increased in the wetlands. This is related to landscape degradation resulting from transformation of forest areas into urban areas, and to soil losses in urban areas due to

**Table 5** Result of Kruskal–Wallis test of water quality parameters within three wetland groups

Water quality parameter	Statistics	
	Adjusted H	<i>P</i> value
pH	0.01	0.99
EC (mS/m)	10.71	0.00*
Turbidity (mg/L)	2.83	0.24
DO (mg/L)	3.94	0.14
TDS (mg/L)	11.19	0.00*
TN (mg/L)	19.90	0.00*
NO <sub>3</sub> <sup>-</sup> (mg/L)	2.49	0.29
DON (mg/L)	9.30	0.01*
NH <sub>4</sub> <sup>+</sup> (mg/L)	7.24	0.03**
NO <sub>2</sub> <sup>-</sup> (mg/L)	7.87	0.02**
DIN (mg/L)	6.23	0.04**

\*Significant at  $p < 0.01$  \*\*Significant at  $p < 0.05$

improper storm water management, which can cause degradation of water quality due to soil erosion and sediment transportation (Kaste et al. 1997; Castillo et al. 2000). This can be expected in study areas that have experienced rapid development in the urban sector. Urban development influences wetlands through vegetation cover clear cutting, drainage, and cut and fill practices. Moreover, following urban development, significant increases can be expected in the volume of surface waters entering wetlands due to the building up of the impervious cover in the wetlands (Water and River Commission 2001). Mouri et al. (2011) cited a significant relationship between the concentration of TN and the area of either urban or agricultural land. Ye et al. (2009) and Norton and Fisher (2000) also reported a negative relationship between forest areas and concentration of TN. Haidary and Nakane (2008) noted that depending on which type of land use is dominant within the catchment of the wetlands, the dominant type of nitrogen might be different in the out-flowing water from the wetlands. They have reported that NH<sub>4</sub><sup>+</sup> was the dominant form of the nitrogen in the wetlands whose catchment is covered by a high percentage of urban area, and (NO<sub>3</sub><sup>-</sup> + NO<sub>2</sub><sup>-</sup>) were sub-dominant in the wetlands whose catchments were covered by a high percentage of agricultural area.

In contrast, the percentage of forest areas in these wetlands has a significant negative relationship with EC ( $r = -0.62$ ,  $p < 0.01$ ), TDS ( $r = -0.68$ ,  $p < 0.01$ ), TN ( $r = -0.68$ ,  $p < 0.01$ ), DON ( $r = -0.43$ ,  $p < 0.05$ ), and NH<sub>4</sub><sup>+</sup> ( $r = -0.55$ ,  $p < 0.01$ ). These relationships suggest that forest areas play a controlling role in regulating the water quality of wetlands. Hence, if the proportion of forest areas increases within the wetlands, annual mean values of EC, TDS, TN, DON, and NH<sub>4</sub><sup>+</sup> would significantly decrease in the wetland water. It has been well documented that there is a positive relationship between watershed land use practices and soil erosion. In particular, the loss of forest cover is associated with increased soil erosion (Cooke and Prepas 1998; Arnold and Gibbons 1996) and diminished water quality (Houlahan and Findley 2004). The findings of this study are in agreement with those of Amiri and Nakane (2009), who found an inverse relationship between water quality parameters such as pH, SS, *E.coli*, TN and TP and the forest area of the catchments of interest. Additionally, the existence of a high percentage of forest area within the catchment of the wetland can affect the role they play as a nutrient sink. Accordingly, Haidary and Nakane (2008) found that those wetlands, whose catchments were covered by a high percentage of forest area, played a sink function for nitrogen in out-flowing water from the wetlands.

Positive relationships were observed between the annual mean of EC ( $r=0.40$ ,  $p<0.05$ ) and TDS ( $r=0.45$ ,  $p<0.05$ ), TN ( $r=0.44$ ,  $p<0.05$ ),  $\text{NH}_4^+$  ( $r=0.56$ ,  $p<0.01$ ) and the proportion of agricultural areas in the watershed of the wetlands. These relationships could originate from the application of chemical fertilizers and soil erosion from farmlands; the fertilizers and eroded soils are carried out by agricultural runoff into the wetland water (EPA 2002). Ukita and Nakanishi (1999) reported that the potential nitrogen loss rates for agricultural areas (5.5–52.5 kg/ha/year) are 1.38 times that of natural vegetation (0.9–38.0 kg/ha/year). However, no significant relationship was observed in this study between the proportion (%) of grassland areas in the watershed of the wetlands and changes in water quality parameters and nitrogen concentrations. Our results are also confirmed by those of Crosbie and Chow-Fraser (1999) and Haidary and Nakane (2008), who found that the wetlands whose catchments were covered by agricultural area tended to be more turbid and nutrient rich compared to those that were located in forest catchments. Amiri and Nakane (2009) reported a direct relationship between the change in the agricultural area of a given catchment and that of suspended solid and total nitrogen.

Three catchment geometrics including catchment area of the wetland ( $r=0.54$ ,  $p<0.01$ ), Strahler order ( $r=0.52$ ,  $p<0.01$ ) and drainage density ( $r=0.53$ ,  $p<0.01$ ), out of five geometric features of the catchments, indicated positive significant relationship with  $\text{NO}_2^-$ . Accordingly, if these geometrics increase, the  $\text{NO}_2^-$  concentration will increase in the wetland. Regarding the Strahler order, a direct significant relationship was observed between this geometric feature of the catchment and the  $\text{NO}_2^-$  concentration in the wetland. It is possible that an increase in the Strahler order is associated with a decline in the general slope of the catchment. In turn, this could occur due to an increase in the agricultural activities in a given catchment.

The result of the Spearman test revealed that a proportion (%) of grey lowland soil has a significant relationship with  $\text{NH}_4^+$  ( $r=0.60$ ,  $p<0.01$ ), that of regosol with DO ( $r=0.42$ ,  $p<0.05$ ) and DON ( $r=-0.44$ ,  $p<0.05$ ); and that of diluvial sand with TDS ( $r=0.41$ ,  $p<0.05$ ) and  $\text{NH}_4^+$  ( $r=0.58$ ,  $p<0.01$ ). Moreover, rhyolite has a positive significant relationship with TN concentration ( $r=0.46$ ,  $p<0.05$ ) and an indirect relationship with DON ( $r=0.49$ ,  $p<0.05$ ) and  $\text{NH}_4^+$  ( $r=0.47$ ,  $p<0.05$ ) in the wetland.

The Kruskal–Wallis test (Table 5) results indicate significant differences in the annual mean of TDS, TN, DON,  $\text{NH}_4^+$ ,  $\text{NO}_2^-$  and DIN within the three wetland groups with  $p<0.05$ . On the other hand, the water quality parameters, namely pH, EC, turbidity,  $\text{NO}_3^-$  and DO did not reveal significant differences in the annual mean between wetland groups with  $p<0.05$ . The analysis of variance for different forms of nitrogen in this study suggest that all forms of nitrogen including nitrite dissolved inorganic nitrogen, dissolved organic nitrogen, and total nitrogen, except nitrate and ammonium, reveal significant differences among wetland groups.

#### 4 Conclusion

Studying the relationships between the proportion of land use types and water quality parameters of wetlands in the 24 wetland sites of this study indicate that the concentration of TDS, TN, DIN, DON,  $\text{NO}_2^-$ , and EC have decreased along with increase in proportion of forest areas within catchment of the wetland. Hence, such the forest component of the landscape within the catchment of the wetlands can be used to regulate the water quality of these water bodies. The findings of this study have also revealed that there were direct relationships between the proportion (%) of urban areas within the catchment of the wetlands

and the annual mean of nutrients such as TN, DON,  $\text{NO}_2^-$ , TDS and EC in the wetland waters. Moreover, an increase in the proportion of agricultural areas increases the annual mean of  $\text{NO}_2^-$  and TN in the wetland sites. Hence, urban and agricultural areas can be considered as water quality ‘disturbing components’ of the landscape. In comparison with the impact of urban area on the water quality, in case of ammonium, of the wetlands, the impact of agricultural areas can be expected to be more significant than that of urban areas. This can be related to the application of chemical fertilizers within agricultural areas. Hence, land use and water resources planning should consider controlling the extent of agricultural and urban areas in wetlands to improve the environmental quality of the wetlands. The results of this study also indicate that degradation of the landscape within wetlands can cause the deterioration of water quality in such water bodies.

**Acknowledgments** The first author (A.H.) would like to thank the Christian Albrecht Universitaet zu Kiel in Germany for their institutional support. The second author (B.J.A) acknowledges the Postdoctoral Fellowship Program of the Alexander von Humboldt Foundation for conducting this research.

## References

- Akratos CS, Tsihrintzis VA, Pechlivanidis I, Sylaios GK, Jerrentrup H (2006) A free-water surface constructed wetland for the treatment of agricultural drainage entering Vassova Lagoon, Kavala, Greece. *Fresenius Environ Bull* 15(12b):1553–1562
- Amiri BJ, Nakane K (2006) Modeling the relationship between land cover and river water quality in the Yamaguchi prefecture of Japan. *J Ecol Field Biol* 29:343–352
- Amiri BJ, Nakane K (2009) Modeling the linkage between river water quality and landscape metrics in the Chugoku district of Japan. *Water Resour Manag* 23:931–956
- Amiri BJ, Sudheer KP, Fohrer N (2012) Linkage between in-stream total phosphorus and land cover in Chugoku district Japan: an ANN approach. *J Hydrol Hydromech* 60:33–44
- APHA (1995) Standard methods for the examination of water and wastewaters, 19th edn. American Public Health Association, Washington
- Arnold CL, Gibbons CJ (1996) Impervious surface coverage: the emergence of a key environmental indicator. *J Am Plann Assoc* 62:243
- Basnyat P, Teeter D, Flynn KM, Lockaby BG (1999) Relationships between landscape characteristics and nonpoint source pollution inputs to coastal estuaries. *Environ Manag* 23:539–549
- Boskidis I, Gikas GD, Pisinaras V, Tsihrintzis VA (2011) Spatial and temporal changes of water quality, and SWAT modeling of Vosvozis River Basin, North Greece. *J Environ Sci Health A* 45(15):421–440
- Brabec E, Schulte S, Richards PL (2002) Impervious surfaces and water quality: a review of current literature and its implications for watershed planning. *J Plan Lit* 16:409–514
- Brooks KN, Folliott PF, Gregersen HM, DeBano LF (2003) Hydrology and the management of watersheds, 3rd edn. Wiley-Blackwell. pp 574
- Castillo MM, Allan JD, Brunzell S (2000) Nutrient concentrations and discharges in a Midwestern agricultural catchment. *J Environ Qual* 29(4):1142–1151
- Clapcott E, Collier KJ, Death G, Goodwin EO, Harding JS, Kelly D, Leathwick JR, Young RG (2011) Quantifying relationships between land-use gradients and structural and functional indicators of stream ecological integrity. *Freshw Biol*. doi:10.1111/j.1365-2427.2011.02696.x
- Cooke SE, Prepas EE (1998) Stream phosphorus and nitrogen export from agricultural and forested watersheds on the boreal plain. *Can J Fish Aquat Sci* 55:2292–2299
- Crosbie B, Chow-Fraser P (1999) Percentage land use in the watershed determines the water and sediment quality of 22 marshes in the Great Lakes basin. *Can J Fish Aquat Sci* 56:1781–1791
- Daley, ML, McDowell WH (2002) Relationship between dissolved organic nitrogen and watershed characteristics in a rural temperate basin. American Geophysical Union, Spring Meeting, Washington D.C., USA
- DeBusk WF (1999) Nitrogen cycling in wetlands, institute of food and agricultural sciences. University of Florida
- Gikas GD, Yiannakopoulou T, Tsihrintzis VA (2006) Water quality trends in a coastal lagoon impacted by non-point source pollution after implementation of protective measures. *Hydrobiologia* 563(1):385–406



- Grant WA, Pederson AK, Marine SL (1997) Ecology and natural resource management: systems analysis and simulation. John Wiley & Sons. Pp. 400
- Gregory MB, Calhoun DL (2007) Physical, chemical, and biological responses of streams to increasing watershed urbanization in the piedmont ecoregion of Georgia and Alabama; Chapter B of effects of urbanization on stream ecosystems in six Metropolitan areas of the United States; U.S. Geological Survey Scientific Investigations Report 2006-5101-B: Reston, VA, USA, 2007; p.104; available online only at <http://pubs.usgs.gov/sir/2006/5101B> (accessed on July 2010)
- Haidary A, Nakane K (2008) Comparative study of the nitrogen dynamics of three wetlands in Higashi-Hiroshima area, west Japan. *Pol J Environ Stud* 18:617–626
- Helms BS, Schoonover JE, Feminella JW (2009) Assessing influences of hydrology, physicochemistry, and habitat on stream fish assemblages across a changing landscape. *J Am Water Resour Assoc* 45:157–169
- Houlahan JE, Findley CS (2004) Estimating the “critical” distance at which adjacent land-use degrades wetland water and sediment quality. *Landsc Ecol* 19:677–690
- Jones KB, Neale AC, Nash MS, Van Remortel RD, Wickham JD, Riitters KH, O’Neill RV (2001) Predicting nutrient and sediment loadings to streams from landscape metrics: a multiple watershed study from the United States mid-Atlantic region. *Landsc Ecol* 16:301–312
- Jordan TE, Whigham DF, Hofmockel KH, Pittek MA (2003) Wetlands and aquatic processes nutrient. *J Environ Qual* 32:1534–1547
- Kaste O, Henriksen A, Hindar A (1997) Retention of atmospherically-derived nitrogen in subcatchments of the Bierkreim river in Southwestern Norway. *Ambio* 26:269–303
- Lee BA, Kwon J, Kim JG (2005) The relationship of vegetation and environmental factors in wangsuk stream and gwarim reservoir: I. Water environments. *Korean J Ecol* 28:365–373
- Liddle MJ (1975) A selective review of the ecological effects of human trampling on natural ecosystems. *Biol Conserv* 7:17–36
- May CW, Horner R, Karr J, Mar B, Welch E (1997) Effects of urbanization on small streams in the puget sound lowland ecoregion. *Watershed Prot Tech* 2(4):483–494
- Mouri G, Takizawa S, Oki T (2011) Spatial and temporal variation in nutrient parameters in stream water in a rural–urban catchment, Shikoku, Japan: effects of land cover and human impact. *J Environ Manage* 92:1837–1848
- Ngoye E, Machiwa J (2004) The influence of land use patterns in the ruvu river catchment on water quality in the river system. *Phys Chem Earth* 29:1161–1166
- Norton MM, Fisher TR (2000) The effects of forest on stream water quality in two coastal plain watersheds of the Chesapeake Bay. *Ecol Eng* 14(4):337–361
- Papastergiadou ES, Retalis A, Apostolakis A, Georgiadis T (2008) Environmental monitoring of spatio-temporal changes using remote sensing and GIS in a Mediterranean Wetland of Northern Greece. *Water Resour Manag* 22:579–594
- Plameri L, Treppel M (2002) A GIS-based score system for siting and sizing of created or restored wetlands: two case studies. *Water Resour Manag* 16:307–328
- Richards C, Host G (1994) Examining land use influences on stream habitats and macro-invertebrates. *Water Resour Bull* 30:729–739
- Schueler T (1994) The importance of imperviousness. *Watershed Prot Tech* 1:100–111
- Shimoda M (1993) Effect of urbanization on pond vegetation in the Saijo basin, Hiroshima prefecture, Japan. *Hikobia* 11:305–312
- Tsihrintzis VA, Hamid R (1997) Modeling and management of urban stormwater runoff quality: a review. *Water Resour Manag* 11(2):137–164
- Tsihrintzis VA, Vasarhelyi GM, Lipa J (1995) Multiobjective approaches in freshwater wetland restoration and design. *Water Int IWRA* 20(2):98–105
- Tsihrintzis VA, Fuentes HR, Gadipudi R (1996) Modeling prevention alternatives for nonpoint source pollution at a wellfield in Florida. *Water Resour Bull AWR* 32(2):317–331
- Tsihrintzis VA, Fuentes HR, Gadipudi R (1997) GIS-aided modeling of nonpoint source pollution impacts on surface and ground waters. *Water Resour Manag* 11(3):207–218
- U.S. EPA (2002) Methods for evaluating wetland condition: land-use characterization for nutrient and sediment risk assessment. Office of Water, U.S. Environmental Protection Agency, Washington, DC. EPA-822-R-02-025
- Ukita M, Nakanishi H (1999) Pollutant load analysis for the environmental management of enclosed sea in Japan. Proceedings of the Fourth International Conference on the Management of Enclosed Coastal Seas
- Water and River Commission (2001) Living wetland: an introduction to wetlands. *Water Facts* 16:1–20
- Ye L, Cai QH, Liu RQ, Cao M (2009) The influence of topography and land use on water quality of Xiangxi River in three Geoges reservoir region. *Environ Geol* 58:937–942

# An Integrated Framework for the Development of Green Infrastructure: A Literature Review

by

P. Beauchamp<sup>1</sup> and J. Adamowski<sup>2</sup>

## **Abstract**

Green infrastructure (GI) has emerged as an active term of reference in project development planning. A gap exists in the GI research literature in the form of the absence of an integrated framework to assist engineering organizations in planning the start-up of new projects in the context of greening and sustainability. This study attempts to identify the existing frameworks that propose the development of green projects.

The first purpose of this study is to explore the use of fully integrated GI in the engineering design of a modern, new development. A clear sequence of tasks must define the workflow, leading teamwork. This literature review identifies several different approaches and selects four to build a ready-to-use framework of sequenced tasks, which includes all the components of water management (rain and drainage, water supply and wastewater).

This essay reviews GI literature with a focus on water resources and ecosystem services. It includes the methods, tools, and techniques available in different approaches, such as low impact development (LID), leadership in energy and environmental design (LEED), the British Columbia (BC) guidebook, and the EPA (Environmental Protection Agency) scorecard. It explores different frameworks to start or structure an urban development project.

*Key Words:* Green Infrastructure, climate change, environment

## **1. Introduction**

While architects and designers are beginning to incorporate biophilia into their work, planners and policymakers who consider cities lag behind; the subject raises serious questions about what a city is or could be, and what constitutes a livable, sustainable environment (Beatly, 2010). Authors like Sim Van Der Ryn promote the concept of ecological design. Perhaps the most compelling theme of ecological design is

<sup>1</sup> P.Eng. Sen. VP International, exp. Services Inc.

<sup>2</sup> Ph.D. Prof. Department of Bioresource Engineering, McGill University

the search for a unified approach to the design of sustainable systems that integrates scales ranging from molecular to global (Van Der Ryn & Cowen, 2007). Many authors propose green infrastructure (GI) as a way to design sustainable city systems. As a generic concept, GI also includes water.

The last ten years have seen extraordinary theoretical and technical advances in the field of ecological design; yet the challenges facing the planet have accelerated, ranging from the loss of biodiversity to rapidly increasing effects of global climate change (Van Der Ryn & Cowen, 2007). Stormwater management is an increasing concern due to climate change and water supply in terms of quality and quantity is also significant. Contamination of water sources is another concern resulting from pollution. These elements can be addressed differently in designing new green developments. The philosophy of integrated water resource management is a concern and must be linked to greening the city. Facing uncertainty, adaptive management (AM) is a suitable approach to define GI.

The concept of AM has been designed primarily to support managers in dealing with highly connected systems (Wietske & Jeffrey, 2005). The development of new metrics (standards of classifying or measuring), techniques (ways of classifying or measuring), and analytical frameworks (perspectives on the utility of classes or measures) is needed (Jeffrey & Geary, 2004). Too often, the green urban agenda forgets the "green," concentrating on energy efficiency and resource management, neglecting the life-enhancing and wonder-expanding dimensions of nature itself (Beatly, 2010). This study intends to contribute and meet the needs of incorporating green concepts in city infrastructure design.

## 2.1 Overview

Greening the city can be categorized and grouped into GI, green projects, green development and green approaches. One objective of this study is to define the start-up of a green development project, establishing an integrated green approach to land development and servicing infrastructure. GI can be defined differently, depending on the context in which it is used. Two main definitions of GI are used; some refer to trees in urban areas as GI because of the "green" benefits they provide, while others use GI to refer to engineered structures (such as water treatment facilities or a green roof) that are designed as environmentally friendly (Benedict & McMahon, 2006).

Here, infrastructure is understood as the substructure system such as aqueducts and pipeline on which the growth of a community depends; while GI may be a resilient landscape that supports a multitude of ecological, economic, and social functions without compromising the sustainability of the resource base (Mell, Roe, & Davies, 2009). GI is the physical environment within and between cities, towns, and villages. It is an interconnected network of open spaces, water bodies and environmental features, and the natural systems that these support (Davies, 2011). When hearing the term "infrastructure," most people think of gray infrastructure such as roads and sewers, or of social infrastructure such as hospitals and schools (Bao, 2010). In

the municipal world, these facilities are termed "built infrastructure." With this definition, GI could be defined as hard infrastructure built respecting green principles.

The built infrastructure design approach needs to be matched to the European GI concept (the concept of hubs and connecting links), maintaining that engineered infrastructures will be designed to support the greening process; this approach is identified as integrated GI. Therefore, for GI to be an integrated concept, it should include both concepts in a single ideology, respecting ecology and mimicking nature. Taking a greener approach to infrastructure development not only mitigates the potential environmental effects of development (e.g. improving stream health and reducing energy use) but makes economic sense also, when considering the influence of conventional development on “natural capital” and the services rendered by natural capital (Olewiler, 2004).

There has been a reluctance to use GI in new developments because it is untested (Alexander & Tomalty, 2002). In 2013, this statement remains accurate. This essay provides a literature review of the available approaches for initiating green development.

## 2.2 General Practices

Most papers in the literature focus on individual aspects of green development and in urban planning. For instance, Tzoulas et al. (2007) formulated a conceptual framework of associations between urban green space, ecosystems, and human health. Benedict and McMahon (2002; 2006) focused on land conservation, defining a vision to face the challenges brought by population growth and proposing to build conservation networks that link land for nature and people. McDonald et al. (2005) also proposed a framework based on a landscape approach. Other examples include Mavsar (2010), who developed the forest component of GI, Amati and Taylor (2010) who studied green belts, and Lehmann (2010), who developed some of the green urbanism principles. Douglas Farr (2008) promoted sustainable planning to support sustainable urbanism in urban design, phasing with nature. Sim Van Der Ryn proposed ecological design (Van Der Ryn & Cowen 2007).

Green infrastructures are typically integrated into the process of urban greening and may follow different approaches to attack urban planning and sustaining green cities. Transportation, clean air, population density, health, and water are topics considered in the philosophy of urbanization. Many of the different trends proposed to develop a concept of urbanization are presented below.

Lehmann (2010) proposed the principles of green urbanism, laid out as a step-by-step manual that can be adjusted for application in various contexts. These principles contain a series of pillars that include energy conservation, the use of new technologies (such as combined heat-and-power or solar cooling), the use of renewable energy sources (such as solar PV, solar thermal, wind on land and offshore, biomass, mini hydro, and geothermal), and the concept of the city of short distances, resulting in multiple benefits for both the environment and the economy (Lehmann, 2010).

Wheeler and Beatly (2009) proposed an approach to sustainable urban development in *The Sustainable Urban Development Reader*, presenting an overview of the field by various authors. The topics covered include land use and urban design, transportation, ecological planning and restoration, energy and materials use, economic development, social and environmental justice, and green architecture and building (Wheeler & Beatly, 2009). The edition asks basic questions: what will our cities and suburban landscapes be like in fifty

and in one hundred years? How can we plan and develop communities that will meet long-term human and environmental needs? The concept of sustainable urban development enables citizens, planners, and policy makers worldwide to explore these questions (Wheeler & Beatly, 2009).

Ritchie and Thomas (2009) published *Sustainable Urban Design: An Environmental Approach*. The book, written with many contributors, identifies major issues in making cities environmentally sustainable. It is vital to move towards sustainability in urban forms, transport, landscape, buildings, energy supply, and all other aspects of city living (Ritchie & Thomas, 2009). Ideas of planning, space and form are a backdrop to many of the points made, but the built environment already suffers from those too confident of their solutions and those who think in "silo"-based terms, over-planning and over-constraining development. The book's contributors believe that an integrated approach is needed (Ritchie & Thomas, 2009).

Beatly and Newman (2009) offer *Green Urbanism Down Under*. Beatly previously wrote *Green Urbanism: Learning from European Cities* (1999), which documents the urban ecology and green urban planning work in 30 European cities. *Green Urbanism Down Under* explores positive stories of innovative practice in Australia. Australia represents a good model to illustrate the adage "Think globally, act locally." Australian cities use a variety of planning instruments—including land use and community plans—to give meaning to sustainability (Beatly & Newman, 2009).

Frumkin, Frank and Jackson (2009) proposed urban planning with a health perspective. *Urban Sprawl and Public Health: Designing, Planning and Building* examines how the built environment affects the population and how building smarter can promote health and well-being and protect the environment. The authors take the approach that both land use and transportation are intrinsic to sprawl, promoting densification of the city to reduce its development footprint.

Birch and Wachter's (2008) *Growing Greener Cities* is a collaboration with different authors, presenting an overview of green and sustainable cities and providing tools for measuring and managing success. The book covers most urban green issues. The authors wrestle with the difficulties of breaking old, anti-greening habits and introducing new practices, detailing successful strategies and practices ranging in scale from regional watershed management to rain barrel placement (Birch & Wachter, 2008).

Van Der Ryn and Cowen (2007) proposed a new concept of ecological design in the first edition of *Ecological Design* published in 1999; a benchmark, pioneering work in eco-design. It is not a design handbook but a quest toward creating a design process that has the preservation and restoration of the ecological commons at its core (Van Der Ryn & Cowen, 2007).

Most of these concepts or approaches do not propose an integrated framework. There is a lack of structure to organize the work of professionals in different specialties from start-up to delivering a final plan. Existing planning and development models do not offer a holistic approach for addressing water issues. Only two of these publications address water as an important element in sustainable development.

Novotny, Ahern and Brown (2010) published *Water Centric Sustainable Communities*, combining landscape, water management, transportation, infrastructure, and triple bottom line assessment into an integrated system, covering best practices in GI and



sustainable development. Reuse is becoming a necessity since a city and its water and waste management cannot be separated from its potable water sources. A city cannot have an unsustainable, adverse effect on downstream users and other cities (Novotny et al., 2010).

Sarté (2010) wrote *Sustainable Infrastructure, The Guide to Green Engineering and Design*, offering an extensive examination of sustainable engineering practices in an urban design context. It addresses the processes and systems of sustainable design for greening infrastructure. It also offers a technical, guided approach to working with water, wastewater, energy, and site design. Creating GI involves designing regenerative systems and establishing new ecologies that thrive (Sarté, 2010). Sarté discusses framework approaches to structure the organization of a project, describing in detail four frameworks: 1) pillars of sustainability; 2) the scale density framework; 3) the transect; and 4) the built form-ecology framework. These frameworks are the matrix used to identify elements covered by a sustainable project.

These two books present an extensive description of techniques to improve sustainability, although neither presents a formal framework for initiating a project. Extending these proposals, the present study proposes to fix water as the central element of any green concept of housing development and offers a start-up framework to initiate a planning process.

In Europe, green practices have previously been introduced in infrastructure design without an over-arching framework. The concept of sustainability has been taken into account together with the ecological aspect in decision-making for urban infrastructure selection. Urban planners and civil engineers have tried to respect the Brundtland Commission's definition of sustainability in designing infrastructure, which states that sustainability is development that meets the needs of the present without compromising the ability of future generations to meet their needs (WCED, 1987).

In 2009, Natural England published this GI guidance:

Green Infrastructure is a strategically planned and delivered network comprising the broadest range of high quality green spaces and other environmental features. It should be designed and managed as a multifunctional resource capable of delivering those ecological services and quality of life benefits required by the communities it serves and needed to underpin sustainability. Its design and management should also respect and enhance the character and distinctiveness of an area with regard to habitats and landscape types (Natural England, 2009, p. 7).

In France, the concept of *trame écologique* was developed from 1990 to 2000. In the following years, numerous regions implemented the concept and in 2010, Allog-Dhuise updated the original document. The term "green infrastructure and blue" was chosen to reflect the new, dual concept, reflecting the importance of provinces and departments expressing their view on the spatialization of this issue. In 2004/5, DIREN Rhône-Alpes, associated with the Loire DDE conducted an experiment in this direction in the territory of South SCoT Loire (Chatain, 2005). France focuses on the broad concept of sustainable development and has adopted a green development strategy. *Trame verte et bleue* includes the concept of hubs and corridors and ecology networks are seen as the key to conserving biodiversity.

In the US, the notion of (low impact development (LID) has been formally structured since 1998, when the Low Impact Development Center (LIDC) was established to design and provide information about new stormwater management techniques. Leadership in energy and environmental design (LEED) was developed in 2002 by the US Green Building Council (USGBC), providing building owners and operators with a concise framework for identifying and implementing practical and measurable green building design, construction, operations, and maintenance solutions (USGBC, 2009).

In Shanghai, GI entails an increase in green spaces in the city. Park departments propose a scientific approach to ensure that all citizens live within two kilometers of a green space (park). An extensive program to develop public parks within the city is currently being implemented.

In Australia, GI presently focuses mainly on green roof and green wall design because many green practices have already been introduced into design practices. Water sensitive urban design (WSUD) has evolved from its earlier association with stormwater management to provide a broader framework for sustainable urban water management, now offering a framework for common and unified methods of integrating the interactions between the urban built form (including urban landscapes) and the urban water cycle (Wong, 2006). GI is the network of designed and natural vegetation found in cities and towns, including public parks, recreation areas, remnant vegetation, residential gardens and street trees, as well as innovative and emerging new urban greening technologies such as the green roof and the green wall (Barlow, 2009).

In 2006, Mitchell explored Australian experiences in the application of the concept of integrated urban water management (IUWM) to land development sites. The understanding of IUWM is maturing within the Australian water industry. Successful projects include the translation of IUWM concepts into well-functioning operational urban developments, significant reductions in the effect of the urban developments on the total water cycle, and the increasing acceptance of the concept within the water and land development industries (Mitchell, 2006). However, there is room for greater integration of the water supply, stormwater, and wastewater components of the urban water cycle, improved dissemination of knowledge, enhancement of skills in both public and private organizations, and in monitoring the performance of systems and technologies (Mitchell, 2006).

In Western Canada, the development of GI is promoted through partnership between the public, government departments, and the private sector. Formal organizations were established in 2006. In Ontario in 2008, the City of Toronto developed the Green Development Standards. Tables that outline the wet weather flow management guidelines for Toronto form one example of the guidelines that make up these standards. The standards also cover the Better Buildings Partnership guidelines. Other terms originating from the same source as LID include sustainable urban drainage systems, innovative or integrated stormwater management, WSUD, GI design, ecological engineering, and the LEED for neighborhood development (LEED ND) (Gyurek, 2009).

There is a multitude of approaches to managing environmental projects. Some authors propose holistic approaches for planning projects. In 2004, John Randolph published *Environmental Land Use Planning and Management*, a textbook that presents a comprehensive

approach to issues of land use planning and management. The author described basic knowledge in planning theory and natural science, focusing on land planning.

In 1996, the EPA (Environmental Protection Agency) entered into a cooperative agreement with the American Society of Civil Engineers led by members of the Urban Water Resources Research Council to initiate the International Stormwater Best Management Practices Database Project (BMP Database). The BMP Database goals were multi-faceted, although key goals include the development of a standardized set of monitoring and reporting protocols for urban stormwater BMP performance studies. The 2012 version of this monitoring manual includes stormwater management practices and a planning approach (WERF, 2012). Again, this approach focuses on a specific topic: stormwater management. The methodology provides an eight-step approach for developing a monitoring plan including defining study objectives, identifying study goals, identifying information inputs/data needs, defining study boundaries, developing an analytical approach, specifying performance or acceptance criteria, developing a detailed plan for obtaining data, and assessing the reasonableness of the plan and refinement.

In 2009, the EPA's Smart Growth Program in conjunction with the Office of Water edited the *Water Quality Scorecard*, incorporating GI practices at the municipal, neighborhood, and site levels. This scorecard offers policy options and a systematic approach for management across multiple municipal departments, proposing design-managing parameters for municipal officers. Again, this approach is oriented to stormwater management.

LID techniques were pioneered by Prince George's County, Maryland, in the early 1990s. The first methodology proposed by the Department of Environment of Maryland is now used by most practitioners and many states in the US. Again, this approach was designed to manage stormwater.

A universal methodology including all aspects of GI and encompassing the different specialties is required. In Canada, the British Columbia (BC), Alberta, Manitoba, and Ontario approaches have been proposed through partnerships with stakeholders. One outcome has been the development of the BC guidebook in 2002, which has since become a reference on GI in Canada (BCWWA, 2010). However, these approaches continue to focus heavily on stormwater management.

Practitioners require short, comprehensive guidelines to plan their projects. LEED is now a reference for green buildings. In 2010, the USGBC developed the LEED ND rating system to guide and assess sustainable community development, the most recent contribution to GI planning. As it is a qualification program, the reference is a rating system, explaining how to guide development and redevelopment projects toward more sustainable design. It is not considered a universal approach. This study investigates these references to determine systemic activities to initiate and develop a new green project.

## 2.3 Concepts and Approaches

### 2.3.1 Specific Frameworks

Leaders and members of professional associations in many disciplines have realized that the current infrastructure and urban planning paradigms have become impediments to achieving sustainable urban development and living (Novotny et al., 2010). Many concepts and approaches have been developed and the most referenced are presented below.

### **2.3.1.1 American Institute of Architects (AIA) Committee on the Environment: Ten Measures of Sustainable Design**

“The Committee on the Environment (COTE) works to advance, disseminate, and advocate design practices that integrate built and natural systems and enhance both the design quality and environmental performance of the built environment”; COTE serves as the community and voice on behalf of AIA architects regarding sustainable design. COTE’s ten measures of sustainable design and performance metrics are: 1) sustainable design intent and innovation, 2) regional/community design and connectivity, 3) land use and site ecology, 4) bioclimatic design, 5) light and air, 6) water cycle, 7) energy flows and energy future, 8) materials and construction, 9) long life, 10) loose fit, and collective wisdom and feedback loops (AIA, 2012).

### **2.3.1.2 American Society of Landscape Architects (ASLA) Sustainable Sites Initiative Benchmarks and Performance Guidelines**

The Sustainable Sites Initiative is an interdisciplinary effort by the American Society of Landscape Architects, the Lady Bird Johnson Wildflower Center at The University of Texas at Austin and the United States Botanic Garden to create voluntary national guidelines and performance benchmarks for sustainable land design, construction and maintenance practices; The Meadows Foundation and Landscape Structures (ASLA, 2009) provide major funding for the Sustainable Sites Initiative. The framework is presented in nine topics: 1) site selection; 2) pre-design assessment and planning; 3) site design—water; 4) site design—soil and vegetation; 5) site design—materials selection; 6) site design—human health and well-being; 7) construction; 8) operations and maintenance; and 9) monitoring and innovation (ASLA, 2009).

### **2.3.1.3 BREEAM**

“BREEAM (building research establishment environmental assessment method) is a building certification system established in 1990. It is a method of environmental auditing, providing a set of standards for best practice in sustainable development for the design, construction, operation and environmental performance of buildings”; The main criteria for calibration include measures affecting energy, water use, indoor environment, pollution, transport, materials, waste, ecology, and management processes (BREEAM, 2012).

### **2.3.1.4 Light Imprint**

LINU (light imprint new urbanism) is a technical development based on the principle of minimum land territory by coordinating the engineering principles of new urbanism

and sustainability, offering a set of solutions through transitional areas. Light imprint principles are formatted in a handbook and the light imprint matrix is the primary organizing method.

"The *Light Imprint Handbook* is a quick reference to which tools can be utilized to best implement light imprint (LI) techniques, also demonstrating where tools are most appropriately located along the transect. The transect zone matrix is designed to serve as an organizational framework and is by its nature somewhat subjective; the LI team suggests where on the rural to urban scale each tool is most useful depending on location on the transect; each project has a specific set of needs" (CNU, 2012).

To create a simple framework, the LI tools are classified into four main categories: paving, channeling, storage and filtration; some tools can be used for more than one function. The LI team's approach is to classify most tools by their principal function and refer to their benefits in other categories; over sixty tools are provided to apply solutions for different applications (CNU, 2012).

#### **2.3.1.5 Living Building Challenge**

"This approach is also a certification program designed on the basis of conservation and restoration and is an integrated tool that can be applied to landscaping, infrastructure, renovations to buildings, campus and community development. The Living Building Challenge is comprised of seven performance areas, or "petals": site, water, energy, health, materials, equity and beauty. Petals are subdivided into twenty imperatives, each focusing on a specific sphere of influence. This compilation of imperatives can be applied to almost every conceivable typology or project type, whether a building (the renovation of an existing structure or new construction), infrastructure, landscape or community development. Naturally, strategies to create living landscapes, infrastructure, renovations, buildings or neighborhoods will vary widely by occupancy, use, construction type and location, but the fundamental considerations remain the same" (ILFI, 2012).

#### **2.3.1.6 Melbourne Principles for Sustainable Cities**

At an international conference in Australia on 2 April 2002 was organized by the United Nations Environment Programme (UNEP) and the International Council for Local Environmental Initiative. Ten principles of sustainable development were outlined to create a comprehensive framework for building better towns.

The Melbourne Principles are intended to guide thinking and provide a strategic framework for action; they are not prescriptive, allowing cities to develop sustainable solutions relevant to their particular circumstances. They can help to bring together citizens and decision-makers, whose participation and cooperation is essential in transforming cities to sustainability. The principles are: 1) provide a long-term vision for cities based on sustainability, intergenerational, social, economic and political equity, and their individuality; 2) achieve long-term economic and social security; 3) recognize the intrinsic value of biodiversity and natural ecosystems, and protect and restore them; 4) enable communities to minimize their ecological footprint; 5) build on the characteristics of ecosystems in the development and nurturing of healthy and sustainable cities; 6) recognize and build on the distinctive characteristics of cities, including their human and



cultural values, history and natural systems; 7) empower people and foster participation; 8) expand and enable cooperative networks to work towards a common, sustainable future; 9) promote sustainable production and consumption through appropriate use of environmentally sound technologies and effective demand management; and 10) enable continual improvement based on accountability, transparency and good governance (UNEP, 2002).

### **2.3.1.7 Net-Zero Energy Development**

"In 2003, a group of homebuilders and developers began informal discussions on new decentralized energy systems and how future Canadian homes could be better designed to respond to Canada's clean air and climate change objectives". In 2006, the coalition was incorporated as a not-for-profit organization to promote energy efficiency in homes. The Net-Zero Energy Home Coalition (NZEH) promotes the development of homes that consume a small amount of energy, focusing its efforts on activities that have the greatest influence on achieving the goals of reducing energy consumption (Net-Zero Energy Home Coalition, 2012). Currently, the coalition extends to North America, creating a large membership including 361 architects.

### **2.3.1.8 One Planet Living's Ten Principles**

"The One Planet Communities program is creating a network of the earth's greenest neighborhoods. One Planet Living is a model based on ten simple principles which provide a framework to make sustainable living easy and affordable for all" (Riddlestone 2013). The One Planet Communities program uses ten guiding principles as a framework to help partners examine the sustainability challenges they face and develop appropriate solutions. These principles were developed as a result of lessons learned from BioRegional's work at the pioneering BedZED eco-village in South London. The ten principles of One Planet Living are: 1) zero carbon; 2) zero waste; 3) sustainability; 4) transporting sustainable materials; 5) local and sustainable food; 6) sustainable water; 7) land use and wildlife; 8) culture and heritage; 9) equity and local economy; and 10) health and happiness. One Planet Living uses ecological footprinting as its key indicator of sustainability (One Planet Living, 2012).

### **2.3.1.9 Permaculture**

"Permaculture is a design system for sustainable development that affects all aspects of the human environment. The system teaches how to build eco-homes, grow food, restore the landscape, restore ecosystems, recover rainwater and build new communities'. This approach has been recognized in more than 20 countries since 1985". There are 12 permaculture design principles: 1) observe and interact; 2) catch and store energy; 3) obtain a yield; 4) apply self-regulation and accept feedback; 5) use and value renewable resources and services; 6) produce no waste; 7) design from patterns of nature; 8) integrate rather than segregate; 9) use small and slow solutions; 10) use and value diversity; 11) use edges and value the marginal; and 12) creatively use and respond to change (Permaculture Institute, 2012).

### **2.3.1.10 Regenerative Development**

“During the late 1970s, John T. Lyle (1934–1998), a Cal Poly Pomona landscape architecture professor, challenged graduate students to envision a community in which daily activities were based on the value of living within the limits of available renewable resources without environmental degradation. Over the next ten years, students and faculty researched the possibilities of creating a community that made use of on-site resources, operated with renewable energy, and worked with biologically based processes” (Lyle Center, 2013). The concept seeks to develop an environment that enhances human activities from a personal residence to a complete district, creating human-made surroundings that provide the setting for human activity, ranging from large-scale civic surroundings to personal spaces (Jenkins, 2009).

#### **2.3.1.11 Rocky Mountain Land Use Institute’s Urban Framework**

“The Rocky Mountain Land Use Institute (RMLUI) Sustainable Community Development Code Framework is sustainable at its core, multidisciplinary in its approach, and contextually oriented. It fully encompasses environmental, economic, and social equity. It is innovative and distinctive by linking natural and manmade systems, incorporating useful features of other zoning systems (e.g. performance and form based), and responds to regional climate, ecology, and culture. The basic organization and approach to each topic is to examine relevant obstacles, incentives, and regulations” (RMLUI, 2009). The Framework incorporates sustainability principles and practices, takes a multi-disciplinary approach, promotes triple bottom-line (environment, economy and social equity), is innovative and distinctive, links natural and human-made systems, incorporates useful features of other zoning systems, responds to regional climate, ecology and culture, identifies relevant obstacles, incentives and regulations (Shutkin & Duerksen, 2011).

#### **2.3.1.12 SmartCode**

“SmartCode is a proposal for a unified development plan for land, presenting a vision codifying zoning, subdivision rules, urban design and architectural options. The code develops a community vision for different avenues of development, taking into account the human habitat of a rural aspect to the urban environment, also applying the principle of transit areas” (CATS, 2012). “The SmartCode is a tool that guides the form of the built environment to create and protect development patterns that are compact, walkable, and mixed use; these traditional neighborhood patterns tend to be stimulating, safe, and ecologically sustainable. The SmartCode requires a mix of uses within walking distance of dwellings so residents are not forced to drive everywhere; it supports a connected network to relieve traffic congestion. Simultaneously, it preserves open lands, as it operates at the regional and community scales” (CATS, 2012).

#### **2.3.1.13 Ascertainment**

Many of these concepts refer to LID techniques and to LEED. LEED has become a measure of acceptance for new buildings and subdivisions in developed countries (Novotny et al., 2010). The International Water Association Panel considers LID as a

green scenario (Novotny et al., 2010). In this study, four main frameworks were identified as more referenced. The next section presents these concepts and their frameworks.

### **2.3.2 The LID Concept**

LID is a low cost, effective alternative to stormwater control technology combining resource conservation in a hydrological functional site design with pollution prevention measures to reduce developmental effects, replicating natural watershed hydrology and water quality. Through a variety of small-scale site design techniques, LID controls runoff discharge, volume, frequency and quality to improve development runoff conditions (LIDC, 2011a).

The LIDC was established in 1998 in the US to promote the use of LID and other sustainable stormwater management techniques. The center's mission is to help communities and institutions address increasingly complex and critical issues associated with their resource protection programs and stormwater management regulations. The organization is a multidisciplinary group of technically skilled professionals seeking to develop new approaches to stormwater management, demonstrate their effectiveness, and assist in integrating them into master planning activities, manuals of practice, and personnel training:

The source control of the LID concept is quite different from conventional treatment (pipe and pond stormwater management site design). Hydrologic functions such as infiltration, frequency, and volume of discharge, and ground water recharge can be maintained with the use of reduced impervious surfaces, functional grading, open channel sections, disconnection of hydrologic flow paths, and the use of bio-retention/filtration landscape areas. LID also incorporates multifunctional site design elements into the stormwater management plan. Such alternative stormwater management practices such as on-lot micro storage, functional landscaping, open drainage swales, reduced imperviousness, flatter grades, increased runoff travel time, and depression storage can be integrated into a multifunctional site design (LIDC, 2011b).

### **2.3.3 The LID Approach**

In the US, many states are developing tools to manage LID programs. For example, in 1999, the Maryland Department of Environmental Resources presented a structured approach, proposing to initiate projects using the five steps listed below.

#### **2.3.3.1 Site Planning**

Fundamental concepts that define the essence of LID technology must be integrated into the site planning process to achieve a successful and workable plan. These concepts include using hydrology as the integrating framework, micromanagement, controlling stormwater at the source, using simplistic, non-structural methods, and creating a multifunctional landscape.

#### **2.3.3.2 Hydrologic Evaluation**

The purpose of the hydrologic evaluation is to determine the level of control required to achieve stormwater management goals for LID sites. The required levels of control may be achieved through the application of various hydrologic tools during the site planning process, the use of integrated management practices (IMP), and supplemental controls. The hydrologic evaluation is performed using hydrologic modeling and analysis techniques. The output of the hydrologic analysis provides the basis for comparison with four evaluation measures (runoff volume, peak runoff, frequency, and water quality control).

#### **2.3.3.3 Integrated Management Practices**

LID IMP are designed for on-lot use. This approach integrates the lot with the natural environment and eliminates the need for large, centralized parcels of land to control end-of-pipe runoff. The challenge of designing a low impact site is that the IMP and site design strategies must provide quantity and quality control and enhancement. This includes ground water recharge through infiltration of runoff into the soil; retention or detention of runoff for permanent storage or for later release; pollutant settling and entrapment by conveying runoff slowly through vegetated swales and buffer strips; and multiple uses of landscaped areas.

#### **2.3.3.4 Erosion and Sediment Control**

Erosion and sediment control and stormwater management are closely interrelated. The application of LID concepts and the associated emphasis on minimizing the areas disturbed and breaking up drainage areas into small, manageable sub-catchment areas is in harmony with the basic principles of erosion and sediment control.

#### **2.3.3.5 Public Outreach Program**

Both the public and developers must be committed to the program and a public consultation process is essential. The LID manual presents a strategy in four steps: 1) define objectives; 2) identify target audiences; 3) develop outreach materials; and 4) distribute outreach materials (Prince George's County 2008). The program can be tailored to specific audiences with a specific message to each audience. Seen as an education program, it can identify several objectives, create marketing tools, promote stewardship to initiate environmental protection measures, show potential cost savings, encourage a sense of community, and ensure proper maintenance measures.

LID is one of the fundamental elements of the framework proposed to initiate new developments. The Maryland Department of Environmental Resources developed a LID framework to initiate a project that considered as one of the baseline scenarios to develop a new framework.

The next section presents LEED, the green scenario preferred by urban development professionals.

#### **2.3.4 The LEED Concept**

LEED is an internationally recognized green building certification system that provides third-party verification that a building or district was designed and built using strategies aimed at improving performance across all metrics (LEED 2012). These metrics include

energy savings, water efficiency, CO<sub>2</sub> emissions reduction, improved indoor environmental quality, and stewardship of resources and sensitivity to their effects (USGBC). There are six categories of certification: (1) new construction; (2) commercial interiors; (3) core and shell; (4) existing buildings; (5) homes; and (6) neighborhood developments. In the US, the USGBC manages the certification program while the Canada Green Building Council (CaGBC) manages the program in Canada. The two organizations are independent, although the US certification may accredit Canadian projects. For example, the Toronto Waterfront Project was certified LEED ND Gold by USGBC.

As LEED is mostly dedicated to buildings, the USGBC has developed the LEED for neighborhood development (LEED ND) rating system to guide and assess sustainable community development. The 2009 LEED ND, for example, is a set of performance standards for certifying the planning and development of new neighborhoods. The intent is to promote healthful, durable, affordable, and environmentally sound practices in building design and construction.

Prerequisites and credits in the rating system address five topics: smart location and linkage, neighborhood pattern and design, GI and buildings, innovation and design process, and regional priority credit. The system was created as a partnership between the USGBC, the Natural Resources Defense Council, and the Congress for the New Urbanism (CNU) and registration opened in April 2010.

The CNU is the leading organization promoting workable, mixed-use neighborhood development, sustainable communities, and healthier living conditions and is one of the major leaders of LEED. The CaGBC has developed the Canadian Alternative Compliance Paths (ACP) for the LEED ND 2009 rating system. The ACP are formally approved approaches that provide clarity and guidance for Canadian projects, addressing sections of the rating system that contain US-specific standards or wording (CaGBC, 2012).

#### **2.3.4.1 The LEED Approach**

Under the LEED approach, projects are accredited using a rating system. After registration, the project design team begins to collect information and perform calculations to satisfy prerequisite and credit documentation requirements. To start a project, LEED concentrates on planning following several basic steps, as outlined here.

##### **2.3.4.1.1 Site Analysis and Programming**

This includes property selection, stakeholder identification and outreach, information gathering, environmental review, conceptual planning and development programming.

##### **2.3.4.1.2 Preliminary Planning**

This includes the initial planning for land use, transportation networks and major facilities, public outreach and the refinement of plans.

##### **2.3.4.1.3 Final Design**

This includes continued public outreach, preparation of the final site plan, infrastructure and building design, and the acquisition of a construction permit. LEED



accreditation will be given if a score of 40 (Certified), 50 (Silver), 60 (Gold), or 80 (Platinum) is obtained. The total number of awardable points is 110, comprising 27 points for smart location and linkage, 44 points for neighborhood pattern and design, 29 points for GI and buildings, six points for the innovation and design process, and four points for regional priority credit. A detailed scorecard is published by the USGBC as a project checklist. The accreditation process follows these steps: registration, smart location, and linkage prerequisite review, conditionally approved plan, pre-certified plan, and certified neighborhood development.

LEED is one of the most widely used standards in the US (Sarté, 2010). While the qualifications of LEED are very developed, initiating a project from these criteria remains tedious. Nevertheless, the proposed framework to manage a LEED project is considered by many urban development professionals as a baseline. The framework proposed by LEED is one of the four approaches used in this paper to develop a formulation of synthesis. One of the weaknesses of LEED is the fact that the framework is focused mainly on new development and is not adapted to brownfield development.

The next section presents the US EPA proposal to formulate a green development project.

### **2.3.5 US Environmental Protection Agency Green Approach**

The US EPA promotes GI development, in particular through publishing a series of documents to support stakeholders interested in introducing green action in projects.

#### **2.3.5.1 Municipal Handbook**

The EPA has developed a *Municipal Handbook* (USEPA, 2012), a series of documents aimed at helping local officials implement GI in their communities. The documents cover specific terms to help municipalities introduce GI in the design of storm management facilities. One chapter identifies and discusses the most common funding options available to communities for funding green stormwater infrastructure, stormwater fees, and loan programs. Another chapter covers street design and various other topics are also discussed. Additionally, the EPA has developed the *Water Quality Scorecard* (US EPA, 2009).

#### **2.3.5.2 Water Quality Scorecard (EPA 231-B-09-001)**

The EPA's *Water Quality Scorecard* was developed to assist local governments in identifying opportunities to remove barriers and revise and create codes, ordinances, and incentives for better water quality protection. It guides municipal staff through a review of relevant local codes and ordinances across multiple municipal departments at the three levels of a local government (municipality, neighborhood, and site) to ensure these codes work together to protect water quality goals.

The two main goals of this tool are to help communities protect water quality by identifying ways to reduce the amount of stormwater flows in a community and to educate stakeholders on the wide range of policies and regulations that affect water quality (US EPA, 2009b). In Canada, the provinces mainly manage their green scenario approaches.

In Quebec, the *Sustainable Development Act* was adopted in 2006, recognizing “the character inseparable from environmental, social and economic activities development.” The law proposes to sustain development through the inclusion of a set of 16 sustainable development principles. Nevertheless, there is no formal green scenario proposed as a framework to organize urban development. Conversely, most provinces orient their sustainable development strategy to climate change actions, adopting a strategy to manage river basins. Green scenario action plans are proposed in Alberta and BC and in Ontario, cities like Toronto have a program. The next section presents Alberta’s LID.

### **2.3.6 The Alberta Low Impact Development Partnership Approach**

The Alberta Low Impact Development Partnership (ALIDP) approach was created in 2004 to address the need to protect and maintain the integrity of the natural environment, while promoting growth, prosperity, and quality of life in Alberta’s communities. The formal creation of the ALIDP Society occurred two years later, in 2006. The ALIDP has a diverse base including municipal and provincial governments, watershed stewardship groups, universities, corporations, and individuals with an interest in promoting LID practices. The Edmonton LID Conference (29 September 2009) covered a series of topics to qualify LID actions. The conference presented examples and lessons learned from different projects implemented mainly in Calgary and Edmonton. From the point of view of project preparation, Van Duin and Gyurek present an approach to LID planning.

Van Duin (2009) presented LID criteria. The rating system addresses environmental, economic, and social issues, allowing developers and their consultants, municipalities and the public to evaluate the relative merits of developments from a watershed protection perspective. The keystone of the matrix to rate a LID project should consider these criteria:

- objectives: no adverse effect on receiving water bodies
- strategies: control pollutant loading
- policy tools: land use bylaws, watershed plans, and master drainage plans
- technology/implementation tools: conduct pollutant loading, remove computation and implement all applicable source control practices.

Gyurek (2009) proposed LID as a multi-barrier approach that uses features at the lot, neighborhood, and watershed level to maintain on-site water balance. A multi-barrier approach at the lot level includes a green roof to reduce or delay runoff, the connection of downspouts to rain gardens and/or storage tanks and cisterns, minimum soil depth criteria, direct runoff to infiltration swales, and the use of harvested rainwater to irrigate vegetation or flush toilets.

A multi-barrier approach at the neighborhood-level may involve reduced road widths, using permeable pavement, the removal of curbs/gutters to direct runoff to swales, the promotion of infiltration box planters, the integration of natural wetlands, and/or the building of constructed wetlands to detain runoff, reduce total loadings, and convey parking lot runoff to swales, and bio-retention areas.

A multi-barrier approach at the watershed level was proposed for the integration of natural wetlands with constructed wetlands for major drainage systems to rehabilitate degraded natural features such as wetlands or creeks, maintain natural stream channels, use wide riparian buffer strips, provide sufficient flooding areas in riparian zones, and to avoid direct discharge even after large rainfall events (Gyurek, 2009). The Alberta conference exposed the LID from a design point of view but exposes no formal framework. Nevertheless, the multi-barrier approach is a concept used to develop this paper's new, integrated concept.

Conversely, BC developed sustainability tools relatively early. Documentation is largely diffused on web sites and Canadian practitioners use many of the proposed guidebooks. The next section summarizes the BC approach.

### 2.3.7 The British Columbia Approach

In BC, the Rainwater Management and Green Infrastructure seminar was initiated by an inter-governmental partnership (IGP) on 11 June 2007. The Water Sustainability Action Plan for British Columbia provides a partnership umbrella for an array of on-the-ground initiatives that promote a "water-centric" approach to community planning and development. One of the tools developed under this umbrella is the water balance model for BC. Developed by an IGP (BC and Fisheries and Ocean Canada) as an extension of *Storm Water Planning: A Guidebook for British Columbia*, the water balance model enables users to visualize ways to implement GI solutions to achieve rainwater runoff source control at the site level. The Water Sustainability Action Plan for British Columbia is sponsored by the province of BC, and its elements are delivered through partnerships.

Under the Action Plan umbrella, the Water Sustainability Committee of the BC Water and Waste Association is the managing partnership and is responsible for providing leadership, facilitation, and organizational services for program delivery. Basic information is provided in a guidebook, refocusing the approach to sustainable ecosystem management. The use of the term "stormwater" suggests there is a problem, whereas "rainwater" is seen as a resource (BCWWA, 2005). The past two decades has seen an evolution to an integrated approach.

The approach described in the guidebook also introduced the concept of performance targets to facilitate implementation of the integrated strategy for managing the complete rainfall spectrum. Rainfall capture means include measures such as rain gardens and infiltration soakaways, runoff controls (which delays overflow runoff by means of detention storage ponds), and flood mitigation (which reduces flooding by providing sufficient hydraulic capacity to "contain and convey") (BCWWA, 2005).

Defining rainfall tiers simply enables a systematic approach to data processing and identification of rainfall patterns, distributions, and frequency. The integrated approach proposed by the guidebook (BCME, 2013) is presented in seven steps:

- (1) secure political interest and support;
- (2) frame the watershed problems and opportunities through a land use working session, drainage working session, ecology working session, and interdisciplinary roundtable session;
- (3) develop objectives and alternative scenarios through flood management scenario modeling and source control scenario modeling;

(4) collect meaningful data and refine scenarios according to concurrent rainfall and stream flow data, data on soils and groundwater, water quality data, and data on fish and their habitats;

(5) evaluate alternatives and develop component plans;

(6) develop an implementation program;

(7) refine through adaptive management.

BC's approach is one of the frameworks used for analysis and some of its proposals should be adopted to develop a new framework.

### 2.3.8 Differentiation

The following table presents the main characteristics of each selected framework. The existing frameworks are prepared to respond to a specific topic. At origin, LID was proposed to solve stormwater management problems, not addressing integrated water resources management issues but only a single component. The *Water Quality Scorecard* addresses a municipality's need to organize the startup of a green project, focusing on institutional organization. LEED is a certification for green projects. The LEED ND addresses an approach to design new development but it does not consider principles of integrated water resource management. The BC guidebook was originally prepared to manage stormwater problems. In its second edition, the guidebook (BCME, 2013) focused on integrated water resource management, but the proposed framework is more a guideline for policy makers and institutional management. The book largely covers stormwater management practices.

These four approaches are presented in the following table alongside their main characteristic.

**Table 1. Differentiation among approaches**

Approaches	Focus	Strength	Weakness
LID	Storm management	Hydraulic analysis	No water supply No wastewater No urban planning
<i>WaterQuality Scorecard</i> (EPA)	Institutional organization	Policy issue	No water supply No wastewater
LEED (USGBC)	Urban planning	Land planning standards	No storm management No water supply No wastewater
BC guidebook	Storm management Institutional policy	Watershed management	No water supply No wastewater

### Conclusion

Most frameworks are designed to satisfy the specific needs of public servants or specific designated professionals like engineers, urban planners, architects or landscape architects. As an example, the MDDEF (Québec Ministry of Sustainable Development,

Environment, Fauna and Parks), adopted a new guide (2011) to manage stormwater, proposing to introduce LID techniques in managing stormwater and pipe sizing. There is no consideration to plan other components of urban planning or other components of infrastructure planning. There is a separate and specific approach to guide design of water supply facilities and a specific approach to design small or large wastewater treatment facilities, but no framework to integrate stormwater, water supply and wastewater with urban planning when creating a new urban development.

Any smart growth development should be planned with this new approach: following nature, water should be a central element of any new concept. There is a strong interaction between each element. Drinking water will become wastewater and rainwater will flow to a lake or a river. Policymaking is not an issue for designers, as it is more a municipal concern. In many cases, policy is already defined, becoming an existing parameter.

Two groups of authors have specifically examined the engineering of GI: in 2010, S. B. Sarté published *Sustainable Infrastructure* and *Water Centric Sustainable Communities* was published by Novotny et al. (2010). Sarté offers several forms of guidance for project planning, creating a unique approach for each project by combining different philosophy or development frameworks. Sarté identified 13 frameworks, the most popular being LEED and BREEAM. Four approaches were identified to analyze sustainable infrastructure and four existing frameworks to organize green projects were suggested:

- Framework 1: pillars of sustainability. This approach presents an analysis based on five elements: water, energy, materials, ecology and community. Analysis of the project is formatted in these terms and it is proposed to proceed with a development evolution of the design following five levels of progression.
- Framework 2: the scale density framework. The approach is defined in four words: water, wastewater, energy, solid waste. The needs analysis is defined according to four levels: the city, the district, the block and the building. The organization becomes a pyramidal structure and presents an overall picture of the final proposal.
- Framework 3: the transect. This approach defines territory into seven areas: T1 (natural), T2 (rural), T3 (suburban), T4 (general urban), T5 (urban center), T6 (urban core), and SD (special district). This approach is a form of territorial organization to establish a balance between each of the zones and to identify needs. The overall plan is determined based on a progression from one area to another by introducing measures of sustainable development.
- Framework 4: the built form-ecology framework. This approach interconnects human actions with natural ecological systems. The method uses drivers to guide development. On the horizontal axis are biodiversity, water, air, land and energy. The vertical axis is divided into habitation/settlement, industry/resource extraction and recreation. The principle consists of establishing an equilibrium balancing all these elements according to the criteria in the appropriate box.

All these approaches or developmental frameworks are elements of reflection appropriate to define a development project and to define a sustainable strategy.



However, none define a formula to initiate a project and carry through to final detailed engineering. These frameworks are rhetorical guidelines to orient development. This paper seeks to create a practical approach to assemble the pillars of a new development, not excluding the use of these rhetorical frameworks to enhance the orientation and define goals.

Secondly, Ahern examines best practices for planning the urban environment in a sustainable manner. Ahern proposes to place water at the center of urban concerns. The concept of ecosystem services now provides a powerful, broadly accepted, logical argument for the protection and responsible development of landscapes justified by the specific functions that landscapes provide, often with direct and measurable economic benefits for human beings (Ahern, 2010). Ahern proposed a six step methodology: 1) ecosystem services (goals and assessments); 2) resilience factors; 3) resilience planning strategies; 4) developing scenarios; 5) urban resilience-sustainability planning; and 6) planning implementation–adaptation. Two elements for which this paper searches are included in this approach: a way to initiate a project and a water-centric approach. Nevertheless, it does not describe what to do or how to do it in detail. The new proposed framework should focus on the startup of new development projects with a water-centric approach.

From the current findings, four approaches were selected to use in developing a new startup framework: Maryland's LID approach was chosen from the US, as it was the first state to develop a strategy to implement LID techniques. The EPA's *Water Quality Scorecard* was selected, as the EPA is the American reference to assess environmental effects. The USGBC's LEED is an internationally recognized green building certification system and for this reason, it was selected as the NGO approach for further examination. Finally, BC's guidebook approach was selected from among the Canadian provinces because BC has been a proactive Canadian province in the implementation of environmental planning since 2000 (Water Sustainability Action Plan for BC), BCME (2010).

The new model should integrate the European concept of natural hubs and links with the American concept of GI. It also extends the stormwater management infrastructure approach with an idea of a looping water cycle, maximizing the reuse of wastewater and stormwater. The fundamental idea is to mimic nature and reuse water. In the context of worldwide soft water shortages (particularly in developing countries), maximizing water conservation, natural storage and water reuse will help to address this problem.

To develop a new urban project, studying and developing the concept in six steps is recommended:

- 1) prepare an inventory to take a picture of the site and understand the stakeholders' needs;
- 2) study hydrology and hydraulic assessment to understand the natural flow of water;
- 3) propose IMP to introduce the new concept of a closed loop;
- 4) develop land planning to mimic nature;
- 5) prepare a consultation to review the stakeholders' needs already identified in Step 1;
- 6) propose a master plan to define an initial solution to design the project.

Future GI research would include developing the proposed framework for initiating new projects. Different areas of research can be analyzed using the proposed

approaches. The main objective of this framework is to integrate the work of many professionals. Other areas of further GI research can support a new conceptual development. Clearly, the introduction of green concept in design will meet the challenges of climate change, which could form another topic of research.

## References

- AIA. (2012). *Committee on the environment (COTE)*. Retrieved from <http://network.aia.org/committeeontheenvironment/home/>
- Allag-Dhuisme, F., Amsallem, J., Barthod, C., Deshayes, M., Graffin, V., Lefevre, C., Salles, E. (coord), Barnetche, C., Brouard-Masson, J., Delaunay, A., Garnier, C.C., & Trouvilliez, J. (2010). *Choix stratégiques de nature à contribuer à la préservation et à la remise en bon état des continuités écologiques—premier document en appui à la mise en oeuvre de la Trame verte et bleue en France*. Proposition issue du comité opérationnel Trame verte et bleue. MEEDDM ed.
- Alexander, D. & Tomalty, R. (2002). *Smart growth and sustainable development: Challenges, solutions and policy directions*. Taylor & Francis. doi: 10.1080/135498302200002757
- Amati, M. & Taylor, L. (2010). From green belts to green infrastructure. *Planning, Practice and Research*, 25(2) 143–155, Taylor & Francis.
- ASLA (American Society of Landscape Architects). (2009). *Guidelines and performance benchmarks*. The Sustainable Sites Initiative. <http://www.asla.org/sites.aspx>
- Bao, L. (2010). *Green infrastructure application in the Chelsea River Subwatershed*. Landscape Architecture & Regional Planning Masters Projects, Paper 4. Amhearst: University of Massachusetts. Retrieved from [http://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1003&context=larp\\_ms\\_projects](http://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1003&context=larp_ms_projects)
- Barlow, S. (2011). *Land and food resources*. Melbourne: The University of Melbourne.
- Barthod, C. & Deshayes, M. (2009). *Trame verte et bleue, the French green and blue infrastructure*. Ministère de l'Écologie, de l'Énergie du Développement durable et de l'Aménagement du territoire. Retrieved from: [www.developpement-durable.gouv](http://www.developpement-durable.gouv)
- BCWWA Water Sustainability Committee, Inter-Governmental Partnership, Convening for Action on Vancouver Island, Water Bucket Website Partnership, Vancouver Island Coordinating Team, and Green Infrastructure Partnership with West Coast Environmental Law. (2010). *Beyond the guidebook 2010: Implementing a new culture for urban watershed protection and restoration in British Columbia*. Retrieved from <http://wcel.org/resources/publication/beyond-guidebook-2010-implementing-new-culture-urban-watershed-protection-and->
- BCME British Columbia Ministry of Environment 2013; *Stormwater Planning: A Guidebook for British Columbia*. <http://www.env.gov.bc.ca/epd/mun-waste/waste-liquid/stormwater/>
- BCME British Columbia Ministry of Environment (2010h): *Sustainability Action Plan for British Columbia*, [http://waterbucket.ca/cfa/category/turning\\_ideas\\_into\\_action/](http://waterbucket.ca/cfa/category/turning_ideas_into_action/)
- Beatly, T. (1999). *Green urbanism: Learning from European cities*. Island Press. Washington
- Beatly, T. & Newman, P. (2009). *Green urbanism down under: Learning from sustainable communities in Australia*. Island Press. Washington
- Beatly, T. (2010). *Biophilic cities: Integrating nature into urban design and planning*. Washington, DC: Island Press. Washington
- Benedict, M. & McMahon, E. (2002). *Green infrastructure: Smart conservation for the 21st century*, Sprawl Watch Clearinghouse, Washington, DC.

- Benedict, M. & McMahon, E. (2006). *Green infrastructure: Linking landscapes and communities*. Island Press, Washington, DC
- Birch, E.L. & Wachter, S.M. (2008). *Growing greener cities: Urban sustainability in the twenty-first century*. Philadelphia: University of Pennsylvania Press.
- BREEAM (Building Research Establishment Environmental Assessment Method). (2012). *About BREEAM*. Retrieved from <http://www.breeam.org/about.jsp?id=66>
- British Columbia Ministry of Environment. (2005). *Stormwater planning guidebook for British Columbia*. Retrieved from <http://www.env.gov.bc.ca/epd/epdpa/mpp/stormwater/stormwater.html>
- British Columbia Inter-Governmental Partnership (BCIGP). (2003). *Water balance model (WBM)*. Retrieved from <http://www.waterbalance.ca>
- CATS (Center for Applied Transect Studies). (2012). *Smart Code version 9.2*. The Town Paper Publisher. Retrieved from <http://www.transect.org/codes.html>
- CaGBC 2012, LEED 2009 for Neighbourhood Development with Canadian Alternative compliance Paths retrieved from <http://www.cagbc.org/Content/NavigationMenu/Programs/LEED/RatingSystems/Neighbourhooddevelopments/default.htm>
- Chatain, M. (2005). *Infrastructures vertes et bleues: guides technique et méthodologique*. Éditeur: DIREN Rhône-Alpes. Retrieved from <http://www.observation-urbaine.certu.equipement.gouv.fr/infrastructures-vertes-et-bleues-a174.html>
- City of Toronto. (2008). *The Toronto green development standards*. <http://www.toronto.ca/planning/environment/index.htm>
- CNU (Congress for New Urbanism). (2012). *Light imprint new urbanism*. Retrieved from <http://www.cnu.org/node/1209>
- Davies, C. (2011). *Stockton Tees green infrastructures strategy*. Retrieved from <http://www.stockton.gov.uk/greeninfrastructure>
- Farr, D. (2008). *Sustainable urbanism: Urban design with nature*. (Foreword by Andres Duany). Wiley, Hoboken, New Jersey
- Frumkin, H., Frank, L., & Jackson, R. (2009). *Urban sprawl and public health: Designing, planning, and building healthy communities*. Island Press, Washington, DC
- Gyurek, L. (2009). *Overview and background of LID, city of Edmonton*. The Edmonton LID Conference (September 29, 2009)
- ILFI (International Living Future Institute). (2012). *Living building challenge*. Retrieved from <https://ilbi.org/lbc>
- Inter-Governmental Partnership and Green Infrastructure Partnership. (2007). *Beyond the guidebook: Context for rainwater management and green infrastructure in British Columbia*. <http://www.waterbucket.ca/cfa/sites/wbccfa/documents/media/403.pdf>
- Hobden, D.W., Laughton, G.E., & Morgan, K.E. (2004). Green space borders: A tangible benefit? Evidence from four neighbourhoods in Surrey BC 1980–2001. *Land Use Policy*, 21(2). <http://dx.doi.org/10.1016/j.landusepol.2003.10.002>
- Jeffrey, P. & Geary, M. (2004). Integrated water resources management: Lost on the road from ambition to realization? In: *WATERMAX Conference*. Beijing, November 2004.
- Jenkins, S. (2009). *Towards regenerative development*. Retrieved from [www.planning.nz](http://www.planning.nz)
- Kloss, G. (2009). *Managing wet weather with green infrastructure: Municipal handbook—Rainwater harvesting policies*. US EPA. Retrieved from <http://www.dep.wv.gov/WWE/Programs/stormwater/MS4/guidance/handbooks/Documents/Rain%20water%20harvesting%20municipal%20handbook.pdf>
- Lehmann, S. (2010). *The principles of green urbanism*. UNESCO chair in Sustainable Urban Development for Asia and the Pacific, Earthscan, London UK.

- LIDC (Low Impact Development Center). (2011a). *Low Impact Development Center homepage*. Retrieved from <http://www.lowimpactdevelopment.org>
- LIDC (Low Impact Development Center). (2011b). *Impact Development Center's sizing tool explained*. Retrieved from [http://www.lid-stormwater.net/bioretenction/bio\\_sizing.htm](http://www.lid-stormwater.net/bioretenction/bio_sizing.htm)
- Lukes & Kloss. (2009). *Managing wet weather with green infrastructure: Municipal handbook—Green streets*. US EPA. Retrieved from [http://water.epa.gov/infrastructure/greeninfrastructure/upload/gi\\_municipalhandbook\\_green\\_streets.pdf](http://water.epa.gov/infrastructure/greeninfrastructure/upload/gi_municipalhandbook_green_streets.pdf)
- Mavsar, R. (2010). *Benefit (value) transfer method*. European Forest Institute. [www.efimed.efi.int](http://www.efimed.efi.int)
- McDonald, L., Allen, W., Benedict, M., & O'Connor, K. (2005). Green infrastructure plan evaluation frameworks. *Journal of Conservation Planning*, 1(1), 12–43.
- MDDEFP (Québec Ministry of Sustainable Development, Environment, Fauna and Parks) (2011) *Guide de gestion des eaux pluviales*, <http://www.mddefp.gouv.qc.ca/eau/pluviales/guide.htm>
- Mell, I.C., Roe, M., & Davies, C. (2009). Exploring the role of green infrastructure in the mitigation of climate change in the urban realm. *Climate Change: Global Risks, Challenges and Decisions*. IOP Publishing IOP Conf. Series: Earth and Environmental Science.
- Mitchell, V.G. (2006). Applying integrated urban water management concepts: A review of Australian experience. *Environmental Management*, 37(5), 589–605. doi: 10.1007/s00267-004-0252-1
- Lyle Center. (2013). *History of the Lyle Center*. Retrieved from <http://www.csupomona.edu/~crs/history.html>
- Natura 2000. (2006). *Networking program*. Retrieved from [www.natura.org](http://www.natura.org)
- Natural England. (2009). *Green infrastructure guidance*. Natural England 2009 catalogue code: NE176, p. 7. Retrieved from [www.naturalengland.org.uk](http://www.naturalengland.org.uk)
- Net-Zero Energy Home Coalition. (2012). *Net-zero energy development*. Retrieved from <http://www.netzeroenergyhome.ca/>
- Neumann, M. (2009). *Village at Griesbach—A case study in LEED®-ND certification*. Canada Lands Company.
- Novotny, V., Ahern, J., & Brown, P. (2010). *Water centric sustainable communities*. John Wiley & Sons.
- Olewiler, N. (2004). *The value of natural capital in settled areas of Canada*. Ducks Unlimited Canada and the Nature Conservancy of Canada.
- One Planet Living. (2012). *One Planet Living's ten principles*. Retrieved from <http://www.oneplanetliving.net/what-is-one-planet-living/the-ten-principles/>
- Pennsylvania Environmental Council. (2009). *Implementing green infrastructure: Developing a winning strategy to fund Philadelphia's ambitious visions*.
- Permaculture Institute. (2012). *Sustainable living, practical learning*. Retrieved from <http://www.permaculture.org/nm/index.php/site/index/>
- Prince George's County Maryland Department of Environmental Resources. (2008). *LID national manuals, programs and planning division (PGDER), with assistance from US EPA*. Retrieved from <http://www.co.pg.md.us/Government/AgencyIndex/DER/ESG/manuals.asp>
- Randolf, J. (2004). *Environmental Land Use Planning and Management*, IslandPress, Washington
- Raoult, P. (2010). *The green and blue infrastructure in mainland France: Challenges and experiences—French environment round table*. Retrieved from [www.legrenelle-environnement.fr](http://www.legrenelle-environnement.fr)
- Riddlestone, S. 2013, *What is One Planet Living*, [www.bioregional.com](http://www.bioregional.com)
- Ritchie, A. & Thomas, R. (2009). *Sustainable urban design: An environmental approach* (2nd ed.). Taylor & Francis.
- RMLUI. (2009). *Sustainable community development code: A code for the 21st century*.

- Rottle, N.D. (2010). *Integrating urban green infrastructure through collaborative visioning: New partners for smart growth*. Seattle, Washington: Green Futures Lab, Department of Landscape Architecture, University of Washington.
- Sarté, S.B. (2010). *Sustainable infrastructure: The guide to green engineering and design*. John Wiley & Sons.
- Shutkin, W. & Duerksen, C. (2011). *Sustainable communities start with sustainable community development codes*. Rocky Mountain Land Use Institute. Retrieved from <http://www.law.du.edu//rmlui>
- Tzoulas, K., Korpela, K., Venn, S., Ylipelkonen, Kazmierczak A., Niemela, & James P. (2007). Promoting ecosystem and human health in urban areas using green infrastructure: A literature review. *Landscape and Urban Planning*.
- UNEP (United Nation Environmental Program). (2002). *Melbourne principles for sustainable cities*. International Environmental Technology Centre. Retrieved from <http://www.unep.org/ietc/>
- US EPA. (1999a). *Low-impact development design strategies: An integrated design approach*. EPA 841-B-00-003.
- US EPA. (1999b). *Low-impact development hydrologic analysis*. EPA 841-B-00-002.
- US EPA. (2009a). *Managing wet weather with green infrastructure*. Washington, DC: United States Environmental Protection Agency.
- US EPA. (2009b). *Water quality scorecard, incorporating green infrastructure practices at the municipal, neighborhood, and site scales*. Retrieved from <http://www.epa.gov/owm/mtb/mtbfact.htm>
- US EPA. (2010). *Managing wet weather with green infrastructure*. Retrieved from [http://cfpub.epa.gov/npdes/home.cfm?program\\_id=298](http://cfpub.epa.gov/npdes/home.cfm?program_id=298)
- USGBC. (2009). *LEED for neighborhood development*.
- USGBC. (2011). *A local government guide to LEED neighborhood development*. Retrieved from [www.usgbc.org/nd](http://www.usgbc.org/nd)
- Van Der Ryn, S. & Cowen, S. (2007). *Ecological design* (10th anniversary ed.) Washington: Island Press.
- Van Duin Bert. (2009a). *LID criteria, city of Calgary*.
- Van Duin Bert. (2009b). *LID vs traditional subdivision study, city of Calgary*.
- WCED. (1987). *Our common future: World Commission on Environment and Development (The Brundtland report)*. Oxford: Oxford University Press.
- WERF, 2012, *International Stormwater BMP Database*, [www.bmpdatabase.org](http://www.bmpdatabase.org)
- Wheeler, S.M. & Beatly, T. (2009). *The sustainable urban development reader* (2nd ed.). Routledge.
- Wietske, M. & Jeffrey, P. (2005). *IWRM and adaptive management: Synergy or conflict?* WB 1 Transition to AM of River Basins; Status NeWater.
- Wong, T.H.F. (2006). An overview of water sensitive urban design practices in Australia: Ecological engineering. *Water Practice & Technology*, 1(1). Prahran, Victoria, Australia: IWA Publishing. doi: 10.2166/WPT.2006018



Received 19.11.2012  
Reviewed 24.01.2013  
Accepted 28.01.2013

A – study design  
B – data collection  
C – statistical analysis  
D – data interpretation  
E – manuscript preparation  
F – literature search

## Drought forecasting using new machine learning methods

Anteneh BELAYNEH<sup>ABCDEF</sup>, Jan ADAMOWSKI<sup>ABCDEF</sup>

Department of Bioresource Engineering, Faculty of Agricultural and Environmental Sciences, McGill University, Quebec, Canada, H9X 3V9; e-mail: [jan.adamowski@mcgill.ca](mailto:jan.adamowski@mcgill.ca)

**For citation:** Belayneh A., Adamowski J. 2013. Drought forecasting using new machine learning methods. Journal of Water and Land Development. No. 18 p. 3–12.

### Abstract

In order to have effective agricultural production the impacts of drought must be mitigated. An important aspect of mitigating the impacts of drought is an effective method of forecasting future drought events. In this study, three methods of forecasting short-term drought for short lead times are explored in the Awash River Basin of Ethiopia. The Standardized Precipitation Index (*SPI*) was the drought index chosen to represent drought in the basin. The following machine learning techniques were explored in this study: artificial neural networks (ANNs), support vector regression (*SVR*), and coupled wavelet-ANNs, which pre-process input data using wavelet analysis (*WA*). The forecast results of all three methods were compared using two performance measures (*RMSE* and  $R^2$ ). The forecast results of this study indicate that the coupled wavelet neural network (*WA-ANN*) models were the most accurate models for forecasting *SPI* 3 (3-month *SPI*) and *SPI* 6 (6-month *SPI*) values over lead times of 1 and 3 months in the Awash River Basin in Ethiopia.

**Key words:** agriculture, artificial neural networks, drought forecast, *SPI*, support vector regression, wavelet analysis

### INTRODUCTION

One of the major challenges of agricultural systems is how to mitigate the impacts of droughts. Droughts impact agricultural systems economically as well as environmentally. With respect to economic impacts, droughts damage agricultural production, and can cause economic damage to industries connected to agricultural production, in addition to causing unemployment as a result of reduced production. From an environmental perspective, droughts can deprive crops and soils of essential precipitation as well as increase the salt content in soils and irrigation systems [MISHRA, SINGH 2010].

To mitigate the impacts of drought an effective and timely monitoring system is required. Effective monitoring of droughts can aid in developing an early

warning system. An objective evaluation of the drought condition in a particular area is the first step for planning water resources in order to prevent and mitigate the impacts of future occurrences of drought. The evaluation and forecasting of drought is made possible by the use of drought indices. There are several drought indices that are commonly used, such as the Palmer Index, the Crop Moisture Index and the Standardized Precipitation Index (*SPI*). The Palmer Index and the *SPI* are traditionally the most popular indices for forecasting drought due to their standardization. For the purposes of comparing drought conditions from different areas, which often have different hydrological balances, the most important characteristic of a drought index is its standardization [BONNACORSO *et al.* 2003]. Standardization of a drought in-

dex ensures independence from geographical position as the index in question is calculated with respect to the average precipitation in the same place [CACCIA-MANI *et al.* 2007].

One of the differences between the Palmer Index and the SPI is that the characteristics of the Palmer Index vary from site to site while those of the SPI do not. Another difference is the Palmer Index has a complex structure with a very long memory, while the SPI is an easily interpreted, simple moving average process [TSAKIRIS, VANGELIS 2004]. This characteristic makes the SPI useful as the primary drought index because it is simple, spatially invariant in its interpretation and probabilistic, allowing it to be used in risk and decision making analysis. The SPI is also more representative of short-term precipitation than the Palmer Index and is thus a better indicator for soil moisture variation and soil wetness [MISHRA, SINGH 2010]. Given the focus of this paper on short-term drought, this characteristic makes the use of the SPI more advantageous. The SPI also provides a better spatial standardization than does the Palmer Index with respect to extreme drought events [LLOYD-HUGHES, SAUNDERS 2002]. The SPI has also been found to be better than the Palmer Index in detecting the onset of a drought event [HAYES 1996]. Given the aforementioned advantages of the SPI over the Palmer Index, the SPI was the drought index chosen to forecast in this study.

The SPI will be forecast using machine learning models in this study, namely artificial neural networks (ANN) and support vector regression models (SVR), respectively. These machine learning or data driven models have become increasingly popular in hydrologic forecasting because they are effective in dealing with the non-linear characteristics of hydrologic data. ANNs have been used to forecast droughts in several studies [BACANLI *et al.* 2008; BARROS, BAWDEN 2008; CUTORE *et al.* 2009; KARAMOUZ *et al.* 2009; MARJ, MEIJERINK 2011; MISHRA, DESAI 2006; MORID *et al.* 2007]. There are several studies where SVRs were used in hydrological forecasting. KHAN and COULIBALY [2006] found that an SVR model was more effective at predicting 3–12 month lake water levels than ANN models. KISI and CIMEN [2009] used SVRs to estimate daily evaporation. Finally, SVRs have been successfully used to predict hourly streamflow [ASEFA *et al.* 2006], precipitation [KISI, CIMEN 2012], and were shown to perform better than ANN models for monthly streamflow prediction [MAITY *et al.* 2010; WANG *et al.* 2009], respectively. However, SVRs have not been directly used in drought forecasting to date.

A major limitation of both ANN and SVR models is their ability to deal with non-stationary data. To overcome this limitation, researchers have increasingly begun to use wavelet analysis to pre-process the inputs of hydrologic data. In this study, wavelet anal-

ysis was used in conjunction with an ANN model and compared to a traditional ANN model and an SVR model. Using these three models, SPI 3 and SPI 6, which are short-term agricultural drought indicators, were forecasted over lead times of 1 and 3 months, respectively. Both SPI 3 and SPI 6 were forecasted in the Awash River Basin of Ethiopia.

## MATERIALS AND METHODS

### THE AWASH RIVER BASIN

In this study, the SPI was forecast in the Awash River Basin of Ethiopia (Fig. 1). Forecasts were made and compared for two stations, the Meisso station and Hirna station, within the basin. These two stations are located within the Middle Awash Basin. The Hirna station is labeled on the map and has an elevation of 308 meters above sea level. The Meisso station is located directly west of the Hirna station at an elevation of 259 meters above sea level. While the study was conducted using all the meteorological stations, the results were presented for these two stations in detail.

Drought is a common occurrence in the Awash River Basin [EDOSSA *et al.* 2010]. Ethiopia's weather and climate are extremely variable both temporally and spatially. The heavy dependence of the population on rain-fed agriculture has made the people and the country's economy extremely vulnerable to the impacts of droughts. Current monthly and seasonal drought forecasts are done using the normalized vegetation index (NDVI). While the NDVI is an effective drought index it is sensitive to changes in vegetation and has limitations in areas where vegetation is minimal. Forecasts of SPI 3 and SPI 6 are not dependent on vegetative cover and can be used as another tool for drought forecasts within the basin and the country as a whole to complement the NDVI forecasts.

The mean annual rainfall of the basin varies from about 1,600 mm in the highlands to 160 mm in the northern point of the basin. The total amount of rainfall also varies greatly from year to year, resulting in severe droughts in some years and flooding in others. The total annual surface runoff in the Awash Basin amounts to some  $4,900 \cdot 10^6 \text{ m}^3$  [EDOSSA *et al.* 2010]. Effective forecasts of the SPI can be used for mitigating the impacts of hydrological drought that manifests as a result of rainfall shortages in the area. The climate of the Awash River Basin varies between a mild temperate climate in the Upper Awash sub-basin to a hot semi-arid climate in both the Middle and Lower sub-basins. The Awash River Basin supports farming, from the growth of lowland crops such as maize and sesame to pastoral farming practices. Rainfall records from 1970–2005 were used to generate SPI 3 and SPI 6 time series from each station. Descriptive statistics for precipitation at the two stations is shown in Table 1.

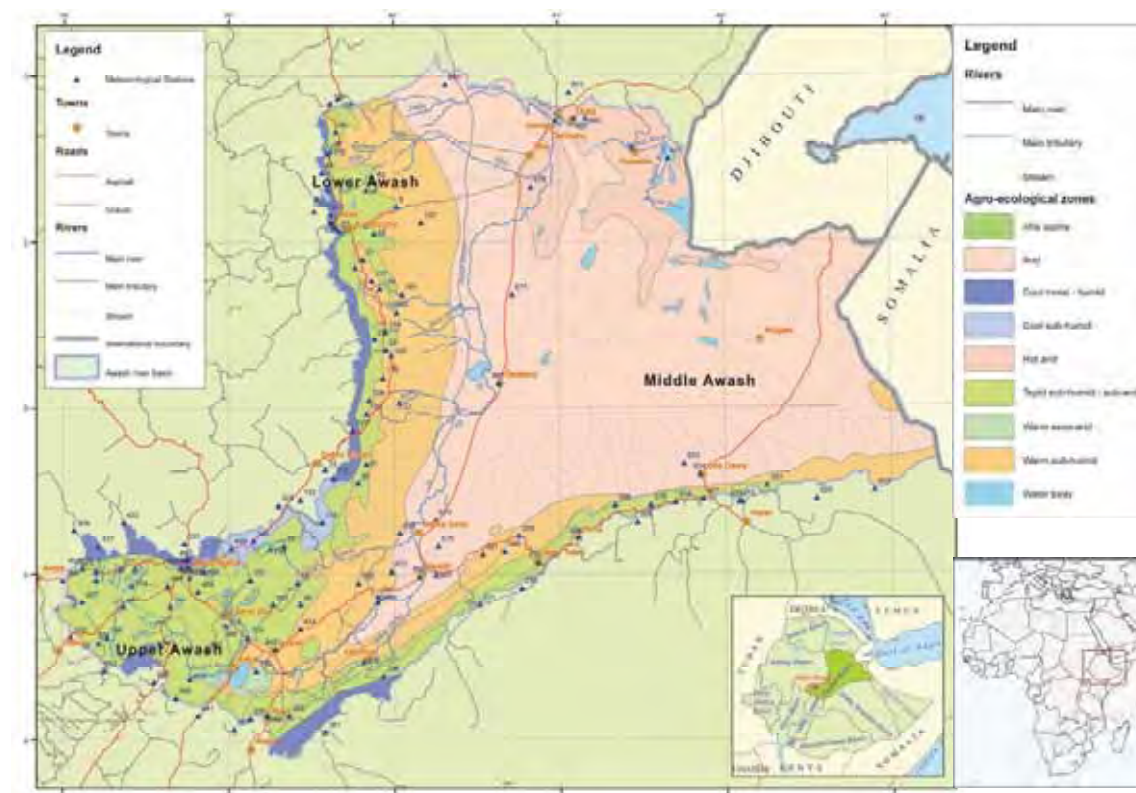


Fig. 1. Awash River Basin; source: Ministry of Water Resources, Ethiopia. Agricultural Water Management Information System [2007]

**Table 1.** Descriptive statistics for stations

Monthly precipitation	Station	
	Meisso	Hirna
Mean	61 mm	78 mm
Max	361 mm	459 mm

#### THE STANDARDIZED PRECIPITATION INDEX (*SPI*)

The standardized Precipitation Index (*SPI*) was developed by MCKEE *et al.* [1993]. The *SPI* index is based on precipitation alone making its evaluation relatively easy compared to other drought indices, namely the Palmer Index and the crop moisture index [CACCIAMANI *et al.* 2007]. In addition to the advantages mentioned earlier, a major advantage of the *SPI* index is that it makes it possible to describe drought on multiple time scales [CACCIAMANI *et al.* 2007, MISHRA, DESAI 2006, TSAKIRIS, VANGELIS 2004]. The *SPI* was also determined to be the best drought index for representing the variability in East African droughts [NTALE GAN 2002], and as such was selected for this study.

The computation of the *SPI* requires fitting a probability distribution to aggregated monthly precipitation series (3, 6, 12, 24, 48 months). The probability density function is then transformed into a normal standardized index whose values classify the cat-

egory of drought characterizing each place and time scale [CACCIAMANI *et al.* 2007]. The *SPI* can only be computed when sufficiently long (at least 30 years) and possibly continuous time-series of monthly precipitation data are available [CACCIAMANI *et al.* 2007]. *SPI* values can be categorized according to classes [CACCIAMANI *et al.* 2007]. Normal conditions are established from the aggregation of two classes:  $-1 < SPI < 0$  (mild drought) and  $0 \leq SPI \leq 1$  (slightly wet). *SPI* values are positive or negative for greater or less than mean precipitation, respectively. Variance from the mean is a probability indication of the severity of the flood or drought that can be used for risk assessment [MORID *et al.* 2007]. The more negative the *SPI* value for a given location, the more severe the drought. The time series of the *SPI* can be used for drought monitoring by setting application-specific thresholds of the *SPI* for defining drought beginning and ending times. Accumulated values of the *SPI* can be used to analyze drought severity. In this study, the *SPI\_SL\_6* program developed by the National Drought Mitigation Centre, University of Nebraska-Lincoln, was used to compute time series of drought indices (*SPI*) for each station in the basin and for each month of the year at different time scales.

A 3-month *SPI* compares the precipitation for that period with the same 3-month period over the historical record. For example, a 3-month *SPI* at the

end of September compares the precipitation total for the July–September period with all the past totals for that same period. A 3-month *SPI* indicates short and medium term trends in precipitation and is still considered to be more sensitive to conditions at this scale than the Palmer Index [HAYES 1996]. A 3-month *SPI* can be very effective in showing seasonal trends in precipitation. The 6-month *SPI* compares the precipitation for that period with the same 6-month period over the historical record. The 6-month *SPI* indicates medium-term trends in precipitation and is still considered to be more sensitive to conditions at this scale than the Palmer Index. A 6-month *SPI* can be very effective in showing the precipitation over distinct seasons [HAYES 1996].

### WAVELET DECOMPOSITION

The wavelet transform is a mathematical tool that provides a time-frequency representation of a signal in the time domain [PARTAL, KISI 2007]. In addition, wavelet analysis can often compress or denoise a signal [KIM, VALDES 2003] and thus, is an effective method for dealing with local discontinuities in a given time series. The continuous wavelet transform (CWT) is defined as the sum over all time of the signal multiplied by scaled and shifted versions of the wavelet function  $\psi$  [NASON, VON SACHS 1999]:

$$W(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} x(t) \psi^* \left( \frac{t - \tau}{s} \right) dt \quad (1)$$

where  $s$  is the scale parameter;  $\tau$  is the translation and \* corresponds to the complex conjugate [KIM, VALDES 2003]. The CWT is not often used for forecasting due to its complexity and long computation times. Instead, the wavelet is discretized in forecasting applications to simplify the numerical calculations. The discrete wavelet transform (DWT) requires less computation time and is simpler to implement [NASON, VON SACHS 1999]:

$$\psi_{j,k}(t) = \frac{1}{\sqrt{|s_0^j|}} \psi \left( \frac{t - k\tau_0 s_0^j}{s_0^j} \right) \quad (2)$$

where  $j$  and  $k$  are integers that control the scale and translation respectively, while  $s_0 > 1$  is a fixed dilation step [CANNAS *et al.* 2006] and  $\tau_0$  is a translation factor that depends on the aforementioned dilation step.

When conducting wavelet analysis, the number of decomposition levels that is appropriate for the data must be carefully selected. In this study, each *SPI* time series was decomposed between 1 and 9 levels. The best results were compared at all decomposition levels to determine the appropriate decomposition

level. The optimal decomposition level varied between models. Once a time series was decomposed into an appropriate level, the subsequent approximation series were either chosen on their own, or they were chosen in combination with the relevant detail series, or the relevant detail series were added together without the approximation series. With most *SPI* time series, choosing just the approximation series resulted in the best forecast results. In some cases, the summation of the approximation series with a decomposed detail series yielded the best forecast results. The appropriate approximation was used as an input to the ANN models.

### ANN MODELS

The ANN models used in this study have a feed forward Multi-layer perceptron (MLP) architecture which was trained with the Levenberg-Marquardt (LM) back propagation algorithm. MLPs have often been used in hydrologic forecasting due to their simplicity. MLPs consist of an input layer, one or more hidden layers, and an output layer [KIM, VALDES 2003]:

$$\hat{y}_k(t) = f_0 \left[ \sum_{j=1}^m w_{kj} \cdot f_n \left( \sum_{i=1}^N w_{ji} x_i(t) + (w_{j0}) + w_{k0} \right) \right] \quad (3)$$

where  $N$  is the number of samples,  $m$  is the number of hidden neurons,  $x_i(t)$  = the  $i^{\text{th}}$  input variable at time step  $t$ ;  $w_{ji}$  = weight that connects the  $i^{\text{th}}$  neuron in the input layer and the  $j^{\text{th}}$  neuron in the hidden layer;  $w_{j0}$  = bias for the  $j^{\text{th}}$  hidden neuron;  $f_n$  = activation function of the hidden neuron;  $w_{kj}$  = weight that connects the  $j^{\text{th}}$  neuron in the hidden layer and  $k^{\text{th}}$  neuron in the output layer;  $w_{k0}$  = bias for the  $k^{\text{th}}$  output neuron;  $f_0$  = activation function for the output neuron; and  $y_k(t)$  is the forecasted  $k^{\text{th}}$  output at time step  $t$  [KIM, VALDES 2003].

The ANN models used to forecast the *SPI* were recursive models. The input data was standardized from 0 to 1. All ANN models, without wavelet decomposed inputs, were created with the MATLAB (R.2010a) ANN toolbox. The hyperbolic tangent sigmoid transfer function was the activation function for the hidden layer, while the activation function for the output layer was a linear function. All the ANN models in this study were trained using the (LM) back propagation algorithm. The LM back propagation algorithm was chosen because of its efficiency and reduced computational time in training models [ADAMOWSKI, CHAN 2011].

There are between 3–5 inputs for each ANN model. The optimal number of input neurons was determined by trial and error, with the number of neurons that exhibited the lowest root mean square error (*RMSE*) value in the training set being selected. The inputs and outputs were standardized between 0 and



1. Traditionally, the number of hidden neurons for ANN models is selected via a trial and error method. However a study by WANAS *et al.* [1998] empirically determined that the best performance of a neural network occurs when the number of hidden neurons is equal to  $\log(N)$ , where  $N$  is the number of training samples. Another study conducted by MISHRA and DESAI (2006) determined that the optimal number of hidden neurons is  $2n + 1$ , where  $n$  is the number of input neurons. In this study, the optimal number of hidden neurons was determined to be between  $\log(N)$  and  $(2n+1)$ . For example, if using the method proposed by WANAS *et al.* [1998] gave a result of 4 hidden neurons and using the method proposed by MISHRA and DESAI [2006] gave 7 hidden neurons, the optimal number of hidden neurons was assumed to be between 4 and 7; thereafter the optimal number was chosen via trial and error. These two methods helped establish an upper and lower bound for the number of hidden neurons. For all the ANN models, 80% of the data was used to train the models, while the remaining 20% of the data was divided into a testing and validation set with each set comprising 10% of the data.

#### WANN MODELS

WANN models were similar to the ANN models used in this study. However, instead of using raw *SPI* data in the input layer, WANN models used *SPI* data that had been decomposed using the ‘a trous’ wavelet transform algorithm. The WANN models used to forecast the *SPI* were also recursive models. The input data was standardized from 0 to 1. The WANN models were trained in the same way as traditional ANN models, with the exception that the inputs were made up from either the approximation series, or a combination of the approximation and detail series after the appropriate wavelet decomposition was selected. There were between 5–7 inputs for each WANN model. The optimal number of input neurons was determined by trial and error, with the number of neurons that exhibited the lowest root mean square error (*RMSE*) value in the training set being selected. The selection process for the optimal number of neurons in the hidden layer was similar to the process used for the ANN models. The inputs and outputs were standardized between 0 and 1. The data was partitioned in exactly the same way as ANN models.

#### SVR MODELS

SVR models adhere to the structural risk minimization principle as opposed to the empirical risk minimization principle used by conventional neural networks [VAPNIK 1995]. As a result, these models reduce the generalization error as opposed to the training error. However, these models take much longer time periods to compute compared to conventional

neural networks. All SVR models were created using the OnlineSVR software created by PARRELLA [2007], which can be used to build support vector machines for regression. The data was partitioned into two sets: a calibration set and a validation set. 90% of the data was partitioned into the calibration set while the final 10% of the data was used as the validation set. Unlike neural networks the data can only be partitioned into two sets with the calibration set being equivalent to the training and testing sets found in neural networks. All inputs and outputs were standardized between 0 and 1.

All SVR models used the nonlinear radial basis function (RBF) kernel. As a result, each SVR model consisted of three parameters that were selected: gamma ( $\gamma$ ), cost ( $C$ ), and epsilon ( $\epsilon$ ). The  $\gamma$  parameter is a constant that reduces the model space and controls the complexity of the solution, while  $C$  is a positive constant that is a capacity control parameter, and  $\epsilon$  is the loss function that describes the regression vector without all the input data [KISI, CIMEN 2011]. These three parameters were selected based on a trial and error procedure. The combination of parameters that produced the lowest *RMSE* values for the calibration data sets were selected.

## RESULTS AND DISCUSSION

In the present study the ability of the aforementioned models to effectively forecast *SPI* 3 and *SPI* 6 over different lead times was evaluated. In the following sections, the forecast results for the best models in the Awash River basin are presented. The forecasts presented are from the validation data sets for time series of *SPI* 3 and *SPI* 6, which are analogous to short-term drought (agricultural drought).

For all the stations used in this study, the forecasts for *SPI* 6 were more accurate than the results for *SPI* 3. In addition, as the forecast lead time is increased from 1 to 3 months the forecast accuracy for all the stations and both *SPI* values decreased. This decrease occurred because as the forecast lead time is increased, the errors between the observed and predicted values accumulate at each lead time.

#### *SPI* 3 FORECASTS

As mentioned earlier the *SPI* forecasts were done and compared for two stations in the Awash River Basin. ANN forecasts for the Meisso station reveal that the forecast results decline considerably as the forecast lead time is increased from 1 month to 3 months. This section presents the results from the validation data sets. The *SPI* 3 forecast results for the Meisso station in terms of  $R^2$  are 0.727 and 0.363 for lead times of 1 and 3 months, respectively. An  $R^2$  value between 0.7–0.9 indicates a high degree of correlation, a value between 0.5–0.7 indicates a moderate



degree of correlation and a value between 0.3–0.5 indicates a low degree of correlation. For forecasts of lead times greater than one month,  $R^2$  values above 0.2 are deemed acceptable [MISHRA *et al.* 2007]. In terms of  $RMSE$  the results were 0.106 and 0.103 for lead times of 1 and 3 months, respectively.  $RMSE$  values closer to zero indicate a high level of precision for models. As the lead time is increased the  $RMSE$  increases slightly, indicating that the forecasts still have a high degree of precision.

The results for SVR models are very similar to the results for ANN models. For forecasts of  $SPI$  3, for both the Meisso and Hirna stations, SVR models have slightly better results than ANN models. For example, for forecasts of 1 month lead time at the Meisso station the best ANN model had results of 0.727 and 0.106 for  $R^2$  and  $RMSE$ , respectively, while the best SVR model had results of 0.716 and 0.086. At the Meisso station, for the forecast of  $SPI$  3 for 3

month lead time, the best SVR model had a noticeably better result in terms of  $R^2$ . The SVR model had an  $R^2$  of 0.498 while the  $R^2$  of the ANN model was 0.363.

Forecasts using WANN models yielded better results. In addition to more accurate results the differences between forecasts of 1 month and 3 months lead time were not as pronounced. Forecast results for 1 month lead time at the Meisso station were 0.96 and 0.029 in terms of  $R^2$  and  $RMSE$ , respectively. When the forecast lead time is increased to 3 months, the results are 0.672 and 0.029, respectively. The forecast results for the Meisso station indicate, that for  $SPI$  3, WANN models provide more accurate forecast results for 1 month lead time and for forecasts of 3 months lead time the results do not decline as much as those for models that do not use wavelet analysis. These trends are further exhibited by the results from the Hirna station as shown in Table 2.

**Table 2.** Forecast results for ANN, SVR and WANN models

Model Type	Meisso				Hirna			
	$SPI$ 3		$SPI$ 6		$SPI$ 3		$SPI$ 6	
	$R^2$	$RMSE$	$R^2$	$RMSE$	$R^2$	$RMSE$	$R^2$	$RMSE$
ANN L1	0.727	0.106	0.835	0.091	0.741	0.108	0.87	0.083
ANN L3	0.363	0.130	0.673	0.130	0.509	0.150	0.724	0.127
SVR L1	0.716	0.086	0.876	0.083	0.738	0.100	0.839	0.084
SVR L3	0.498	0.110	0.592	0.110	0.520	0.122	0.612	0.116
WANN L1	0.962	0.029	0.981	0.025	0.982	0.023	0.996	0.012
WANN L3	0.674	0.029	0.797	0.027	0.789	0.089	0.872	0.067

Explanations: L1 – 1-month forecast lead time; L3 – 3-month forecast lead time.  
Source: own results.

### $SPI$ 6 FORECASTS

Forecast results for  $SPI$  6 followed a similar pattern to the results for  $SPI$  3. For the Meisso station, ANN models had forecast results of 0.835 and 0.091 in terms of  $R^2$  and  $RMSE$ , respectively. When the forecast lead time was increased from 1 to 3 months the forecast results were 0.673 and 0.13 in terms of  $R^2$  and  $RMSE$ , respectively. WANN models had more accurate results for both lead times at the Meisso station as shown in Table 2. For forecasts of 1 month lead time, the best WANN model had results of 0.981 and 0.025 in terms of  $R^2$  and  $RMSE$ , respectively. When the forecast lead time was increased to 3 months, the results deteriorated to 0.797 and 0.027, respectively.

For forecasts of 1 month lead time, SVR models have very similar results to those of ANN models for both stations. However, when the forecast lead time is increased to 3 months, it seems that ANN models have slightly better results than those of SVR models. These results are presented in Table 2. An example is at the Hirna station where a  $SPI$  6 forecast of 3 months lead time using the best SVR model yields

results of 0.612 and 0.116 in terms of  $R^2$  and  $RMSE$ , respectively, while the best ANN model has results of 0.724 and 0.127.

For a 1 month lead time, WANN models had results of 0.98 and 0.025 in terms of  $R^2$  and  $RMSE$ , respectively. When the forecast lead time is increased to 3 months, the forecast results do not decrease as much as ANN models. For 3 months lead time, WANN models had forecast results of 0.797 and 0.027 in terms of  $R^2$  and  $RMSE$ , respectively. The forecast results at the Hirna station follow the same pattern exhibited at the Meisso station. The results for the Hirna station, along with the model configurations are presented in Table 2.

### DISCUSSION

The results indicate that ANN, SVR and WANN models can be used as a means of forecasting the  $SPI$  at 1 and 3 month lead times in the Awash River Basin. Figures 2–5 show the forecast results for ANN and WANN models. The figures illustrate how the predicted values closely mirror the observed  $SPI$  values. The results indicate that the use of wavelet analy-

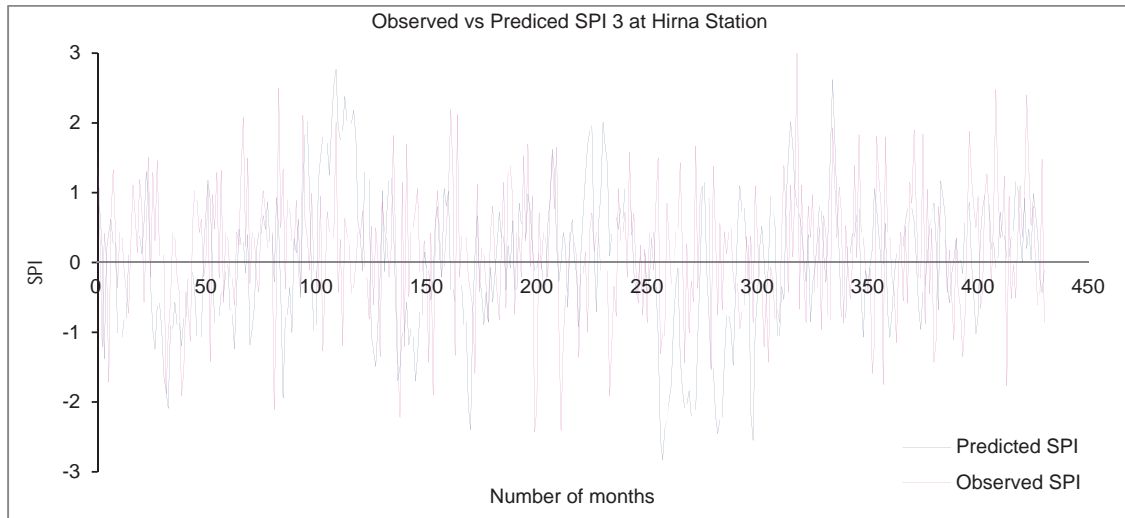


Fig. 2. *SPI 3* forecasts using ANN model at the Hirna station; source: own results

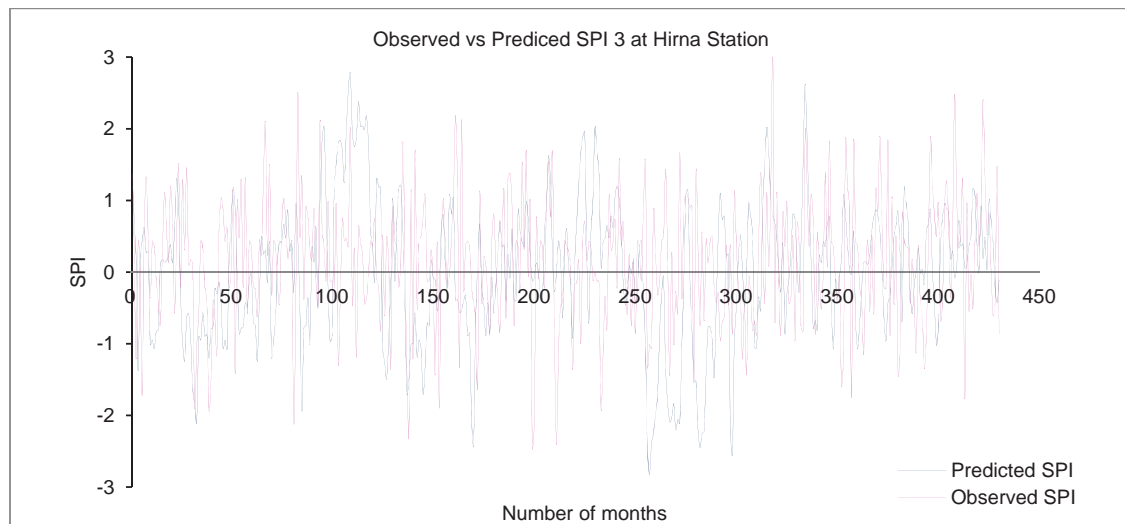


Fig. 3. *SPI 3* forecasts using WANN models for the Hirna station; source: own results

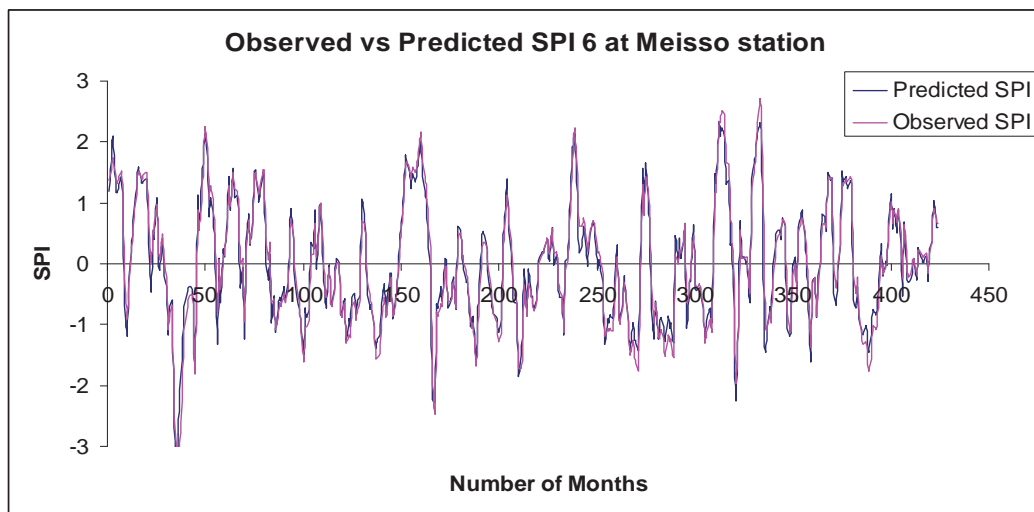


Fig. 4. *SPI 6* forecasts using ANN models at the Meisso station; source: own results

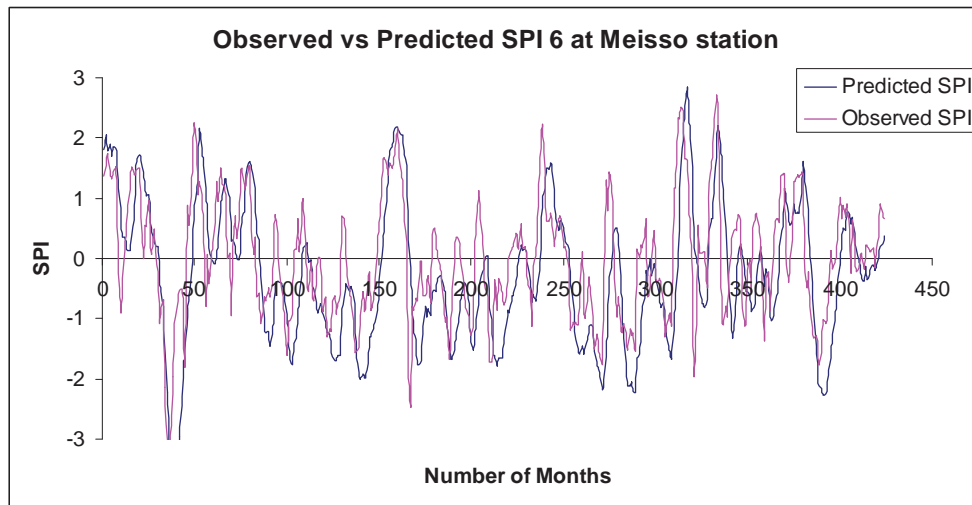


Fig. 5. *SPI* 6 forecasts using WANN models at the Meisso station; source: own results

sis as a pre-processing tool provided good forecast results for both ANN and SVR models irrespective of forecast lead time. As might be expected, the results also indicate that as the forecast lead time is increased the correlation between observed and predicted values as measured by  $R^2$ , decreases and the *RMSE* between observed and predicted values increases.

The results from all the models indicate that *SPI* 6 forecasts were more accurate than *SPI* 3 forecasts. As shown in Table 1, the results for *SPI* 6 are more accurate across all model types. Figure 4 and 5 show how the predicted *SPI* 6 values mirror those of the observed values very closely. Both *SPI* 6 and *SPI* 3 are short-term *SPI* and each new month has less impact on the period of sum precipitation [MCKEE *et al.* 1993]. As a result, monthly variation in precipitation has a smaller impact for *SPI* 6 compared to *SPI* 3. Since *SPI* 6 is a longer term *SPI* its sensitivity to changes in precipitation is less than that of *SPI* 3. This lack of sensitivity may explain why the forecast results for *SPI* 6 are more accurate than those of *SPI* 3. Furthermore, the use of wavelet analysis seems to have improved the results for *SPI* 3 compared to those of *SPI* 6. The inherent sensitivity that shorter term *SPI* values have to changes in precipitation within the precipitation record is why their results are not as accurate as longer term *SPI*. The use of wavelet analysis appears to reduce this sensitivity and while it improves forecasts for both *SPI* 3 and *SPI* 6, it is more beneficial for forecasts of *SPI* 3.

## CONCLUSION

The ability of machine learning techniques and wavelet transforms to forecast the *SPI* drought index over 1 and 3 month lead times was investigated in this study. The results indicate that including wavelet analysis to pre-process data is an effective drought

forecasting tool and improved the results of traditional ANN models. Overall, coupled wavelet-neural network (WANN) models were found to provide the best results for forecasts of *SPI* 3 and *SPI* 6 in the Awash River Basin, especially for *SPI* 6. WANN models showed a higher coefficient of determination between observed and predicted *SPI* compared to ANN and SVR models. WANN models also consistently showed lower values of *RMSE*. It is likely that the coupled models provide accurate results because pre-processing the original *SPI* time series with wavelet decompositions ‘de-noises’ the *SPI* time series, which subsequently allows the ANN models to model the main signal without the noise. Wavelet analysis also seems to reduce the sensitivity to changes in monthly precipitation within the *SPI* time series especially for *SPI* 3 compared to *SPI* 6. This reduction in sensitivity would be more pronounced in shorter-term *SPI* as they are inherently more sensitive to changes in monthly precipitation than longer-term *SPI*.

## ACKNOWLEDGEMENTS

An NSERC Discovery and FQRNT New Researcher Grant held by Dr. Jan Adamowski were used to fund this research. The data was obtained from the Meteorological Society of Ethiopia (NMSA). Their help is greatly appreciated.

## REFERENCES

- ADAMOWSKI J., CHAN H.F. 2011. A Wavelet Neural Network Conjunction Model for Groundwater Level Forecasting. *Journal of Hydrology*. Vol. 407. Iss. 1–4 p. 28–40.
- ASEFA T., KEMBLOWSKI M., MCKEE M., KHALIL A. 2006. Multi-time scale stream flow 505 predicitions: The support vector machines approach. *Journal of Hydrology*. Vol. 318. Iss. 1–4 p. 7–16.
- BACANLI U.G., FIRAT M., DIKBAS F. 2008. Adaptive Neuro-Fuzzy Inference System for drought forecasting. *Stochastic Environmental Research and Risk Assessment*. Vol. 23. Iss. 8 p. 1143–1154.

- BONACCORSO B., BORDI I., CANCELLIERE A., ROSSI G., SUTERA A. 2003. Spatial variability of drought: an analysis of *SPI* in Sicily. *Water Resources Management*. Vol. 17. Iss. 4 p. 273–296.
- BARROS A., BOWDEN G. 2008. Toward long-lead operational forecasts of drought: An experimental study in the Murray-Darling River Basin. *Journal of Hydrology*. Vol. 357. Iss. 3–4 p. 349–367.
- CACCIAMANI C., MORGILLO A., MARCHESI S., PAVAN V. 2007. Monitoring and forecasting drought on a regional scale: Emilia-Romagna Region. In: *Methods and tools for drought Analysis and management*. Eds. G. Rossi, T. Vega, B. Bonaccorso. Water Science and Technology Library. Vol. 62 (1) p. 29–48.
- CANNAS B., FANNI A., SIAS G., TRONCI S., ZEDDA M.K. 2006. River flow forecasting using neural networks and wavelet analysis. *Proceedings of the European Geosciences Union*.
- CUTORE P., DI MAURO G., CANCELLIERE A. 2009. Forecasting Palmer Index using neural networks and climatic indexes. *Journal of Hydrologic Engineering*. Vol. 14. No. 6 p. 588–595.
- EDOSSA D.C., BABEL M.S., GUPTA A.D. 2010. Drought analysis on the Awash River Basin, Ethiopia. *Water Resource Management*. Vol. 24. Iss. 7 p. 1441–1460.
- HAYES M. 1996. Drought indexes. National Drought Mitigation Center, University of Nebraska–Lincoln, p. 7 (available from University of Nebraska–Lincoln, 239LW Chase Hall, Lincoln, NE 68583).
- KARAMOUZ M., RASOULI K., NAZIL S. 2009. Development of a Hybrid Index for Drought Prediction: Case study. *Journal of Hydrologic Engineering*. Vol. 14. No. 6 p. 617–627.
- KHAN M.S., COULIBALY P. 2006. Application of support vector machine in lake water level prediction. *Journal of Hydrologic Engineering*. Vol. 11. No. 3 p. 199–205.
- KIM T., VALDES J.B. 2003. Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. *Journal of Hydrologic Engineering*. Vol. 8. No. 6 p. 319–328.
- KISI O., CIMEN M. 2009. Evapotranspiration modelling using support vector machines. *Hydrological Science Journal*. Vol. 54. Iss. 5 p. 918–928.
- KISI O., CIMEN M. 2011. A wavelet-support vector machine conjunction model for monthly streamflow forecasting. *Journal of Hydrology*. Vol. 399. Iss. 1–2 p. 132–140.
- MAITY R., BHAGWAT P.P., BHATNAGAR A. 2010. Potential of support vector regression for prediction of monthly streamflow using endogenous property. *Hydrological Processes*. Vol. 24. Iss. 7 p. 917–923.
- MARJ A.F., MEIJERINK A.M. 2011. Agricultural drought forecasting using satellite images, climate indices and artificial neural network. *International Journal of Remote Sensing*. Vol. 32. No. 24 p. 9707–9719.
- MCKEE T.B., DOESKEN N.J., KLEIST J. 1993. The relationship of drought frequency and duration to time scales. 8th Conference on Applied Climatology. 17–22 January 1993, Anaheim, California. American Meteorological Society, Anaheim, CA.
- Ministry of Water Resources. 2007. The Awash River Basin. [http://www.mowr.gov.et/AWMISSET/images/Awash\\_agroecologyv3.pdf](http://www.mowr.gov.et/AWMISSET/images/Awash_agroecologyv3.pdf). Accessed 06-June-2013
- MISHRA A.K., SINGH V.P. 2010. A review of drought concepts. *Journal of Hydrology*. Vol. 391. Iss. 1–2 p. 202–216.
- MISHRA A.K., DESAI V.R. 2006. Drought forecasting using feed-forward recursive neural network. *Ecological Modelling*. Vol. 198. No. 1–2 p. 127–138.
- MISHRA A.K., DESAI V.R., SINGH V.P. 2007. Drought forecasting using a hybrid stochastic and neural network model. *Journal of Hydrologic Engineering*. Vol. 12. Iss. 6 p. 626–638.
- MORID S., SMAKHTIN V., BAGHERZADEH K. 2007. Drought forecasting using artificial neural networks and time series of drought indices. *International Journal of Climatology*. Vol. 27. Iss. 15 p. 2103–2111.
- NASON G.P., VON SACHS R. 1999. Wavelets in time-series analysis. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. Vol. 357. No. 1760 p. 2511–2526.
- NTALE H.K., GAN T.Y. 2003. Drought indices and their application to East Africa. *International Journal of Climatology*. Vol. 23. Iss. 11 p. 1335–1357.
- PARRELLA F. 2007. Online support vector regression. Master Thesis, University of Genoa.
- PARTAL T., KISI O. 2007. Wavelet and neuro-fuzzy conjunction model for precipitation forecasting. *Journal of Hydrology*. Vol. 342. Iss. 1–2 p. 199–212.
- TIWARI M.K., CHATTERJEE C. 2010. Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach. *Journal of Hydrology*. Vol. 394. Iss. 3–4 p. 458–470.
- TSAKIRIS G., VANGELIS H. 2004. Towards a Drought Watch System based on spatial SPI. *Water Resources Management*. Vol. 18. Iss. 1 p. 1–12.
- VAPNIK V. 1995. *The nature of statistical learning theory*. New York, USA. Springer Verl. ISBN-10: 0387987800 pp. 334.
- WANG W.C., CHAU K.W., CHENG C.T., QIU L. 2009. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology*. Vol. 374. Iss. 3–4 p. 294–306.
- WANAS N., AUDA G., KAMEL M.S., KARRAY F. 1998. On the optimal number of hidden nodes in a neural network. In: *Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering*. Michigan. IEEE p. 918–921.

**Anteneh BELAYNEH, Jan ADAMOWSKI**

**Prognozowanie suszy z wykorzystaniem automatycznych samouczących się metod**

STRESZCZENIE

**Słowa kluczowe:** *analiza falowa, prognoza suszy, regresja wektorowa, rolnictwo, SPI, sztuczna sieć neuronowa*

Efektywna gospodarka rolna, uzyskanie dużych plonów wymaga prowadzenia działań w celu ograniczenia niekorzystnego wpływu suszy. Ważnym czynnikiem ograniczania skutków suszy jest efektywna i możliwie precyzyjna metoda przewidywania suszy. W artykule przedstawiono trzy metody prognozowania suszy w okresach krótkoterminowych, które zostały zastosowane w zlewni rzeki Awash w Etiopii. Do kwantyfikacji suszy zastosowano wskaźnik standaryzowanego opadu (*SPI*). Zastosowane zostały następujące samouczące się metody: sztuczne sieci neuronowe (ANNs), regresje wektorowe (SVR) oraz połączenie ANNs z analizą falową (WA), którą zastosowano do wstępnej obróbki danych. Ocenę prognozą dokonano stosując dwa mierniki – *RMSE* i  $R^2$ . Na podstawie obliczeń stwierdzono, że połączona metoda analizy falowej z siecią neuronową (WA-ANN) jest najdokładniejsza w prognozowaniu wartości *SPI* 3 (3-miesięczne *SPI*) i *SPI* 6 (6-miesięczne *SPI*) w okresie 1 i 3 miesiące naprzód w zlewni rzeki Awash.



# Determining the amplitude and timing of streamflow discontinuities: A cross wavelet analysis approach

Jan Adamowski<sup>1\*</sup> and Andreas Prokoph<sup>2,3</sup>

<sup>1</sup> Department of Bioresource Engineering, McGill University, 21 111 Lakeshore Road, Ste. Anne de Bellevue, QC, Canada, H9X 3V9

<sup>2</sup> Department of Earth Sciences, University of Ottawa, Marion Hall, Ottawa, ON, Canada, K1N 6N5

<sup>3</sup> Speedstat, Ottawa, ON, Canada, K1G 5J5

## Abstract:

Records of natural processes, such as gradual streamflow fluctuations, are commonly interrupted by long or short disruptions from natural non-linear responses to gradual changes, such as from river-ice break-ups, freezing as a result of annual solar cycles, or human causes, such as flow blocking by dams and other means, instrument calibrations and failure. The resulting abrupt or gradual shifts and missing data are considered to be discontinuities with respect to the normal signal. They differ from random noise as they do not follow any fixed distribution over time and, hence, cannot be eliminated by filtering. The multi-scale resolution features of continuous wavelet analysis and cross wavelet analysis were used in this study to determine the amplitude and timing of such streamflow discontinuities for specific wavebands. The cross wavelet based method was able to detect the strength and timing of abrupt shifts to new streamflow levels, gaps in data records longer than the waveband of interest and a sinusoidal discontinuity curve following an underlying modeled annual signal at  $\pm 0.5$  year uncertainty. Parameter testing of the time-frequency resolution demonstrated that high temporal resolution using narrow analysis windows is favorable to high-frequency resolution for detection of waveband-related discontinuities. Discontinuity analysis on observed daily streamflow records from Canadian rivers showed the following: (i) that there is at least one discontinuity/year related to the annual spring flood in each record studied, and (ii) neighboring streamflows have similar discontinuity patterns. In addition, the discontinuity density of the Canadian streamflows studied in this paper exhibit 11-year cycles that are inversely correlated with the solar intensity cycle. This suggests that more streamflow discontinuities, such as through fast freezing, snowmelt, or ice break-up, may occur during years with slightly lowered solar insolation. Copyright © 2013 John Wiley & Sons, Ltd.

KEY WORDS streamflow discontinuity; continuous wavelet analysis; cross wavelet analysis; solar cycle; Canada

Received 24 July 2012; Accepted 25 March 2013

## INTRODUCTION

Analysis of time series data is an important component in understanding hydrological systems. Discontinuities in normally persistent quasi-stationary streamflow signals, such as annual flow cycles, can provide insight into significant disruptions of the normal flow pattern. For example, gradual changes in forcing mechanisms, such as from the Earth's annual rotation around the sun can lead to non-linear responses such as abrupt freezing, ice break-up, snow melt, or flash floods from lake outbursts. These non-linear responses may not result in extreme flow magnitudes but in dilation or compression of the normal temporal flow pattern that form difficult to detect discontinuities. Other examples of discontinuities are those due to dam opening/closing, spill overflows,

saturation of wetlands or the closing of tributaries. The determination of such discontinuities can assist in understanding annual flow regimes, as well as contribute to identifying trends and forecasting events such as droughts. However, discontinuities in the long-term pattern can be hidden by higher frequency signals, such as noisy streamflow response to heavy rainfall and other short-term meteorological events or by missing or incorrect measurements.

Statistical methods of detecting abrupt changes in streamflow data include the segmented regression with constraints method and Pettitt's test (Pettitt, 1979). The segmented regression with constraints method allows for the detection of both shifts in trend and abrupt changes in hydrological data. Shao *et al.* (2010) applied the technique to analyse shift trends and change points in mean annual temperature, precipitation and runoff time series data in the Shiyang River Basin. Pettitt's test is a non-parametric statistical test that detects a significant change in the distribution of a time series, which may hint towards the existence of discontinuities. Although this

\*Correspondence to: Jan Adamowski, Department of Bioresource Engineering, McGill University, 21 111 Lakeshore Road, Ste. Anne de Bellevue, QC, Canada, H9X 3V9  
Email: jan.adamowski@mcgill.ca

method is able to detect such a change, it does not provide insight into the nature of the change (Zhao *et al.*, 2011). A derivative of the Mann–Kendall test has been used to detect abrupt changes in climate records that can be also readily applied to streamflow records (Goossens and Berger, 1987). Fourier spectral analyses have also been used to study periodicities in streamflow time series data (Kunhel *et al.*, 1990). However, Fourier analysis does not allow for the localization of a particular event in time and is not appropriate for non-stationary processes (Smith *et al.*, 1998). These shortcomings are addressed by windowed short-term Fourier transform methods, such as the Gabor transform, although these methods do not allow for high resolution in both time and frequency.

The wavelet transform overcomes the problems of the methods described above and, thus, has been used as a time-series analysis method for detecting transient events in non-stationary processes in many applications. It transforms a time series into a frequency domain by simultaneously transforming the ‘depth’ or ‘time’ domain and the ‘scale’ or ‘frequency’ domain using various shapes and sizes of short filtering functions called wavelets. Jiang *et al.* (2002) introduced a method based on the Student’s *t*-test to detect the coherency of abrupt changes between two time series, including one hydrological record. Their methods also provided significance estimates for the magnitude of the changes, and they also implemented a feature of the Haar wavelet transform for multi (frequency) separation of these changes. However, the method by Jiang *et al.* (2002) requires two time series; thus, it cannot distinguish the source (time series 1 or 2) of noise-related patterns.

The continuous wavelet transform (CWT) allows for the automatic localization of periodic signals, gradual shifts, abrupt interruptions, trends and onsets of trends in time series (Rioul and Vetterli, 1991; Mallat and Hwang, 1992). Similarly, the cross-wavelet transform allows for the automatic localization of periodic signals, gradual shifts, abrupt interruptions, trends and onsets of trends between two time series (e.g. Maraun and Kurths, 2004; Labat, 2010). As the wavelet function can be rescaled over a variety of timescales (Smith *et al.*, 1998), abrupt changes or discontinuities in signal patterns can be located in terms of both time and frequency (Labat *et al.*, 2000). Therefore, the onset time, duration and dominant frequencies of disturbances can be efficiently extracted with this technique. For example, an investigation of the capabilities of wavelet transforms for detecting signal disturbances in electrical power supply found the modulus maxima of wavelet transforms to be useful in isolating transient disturbances from the rest of the signal. However, in this study, only high-frequency disturbances were investigated, and so the efficiency of wavelet transforms for detecting lower-frequency or superimposed disturbances that are common in hydro-

logical records (e.g. due to waterway blocking from dams or ice) was not determined (Nguyen and Hoang, 1999).

The multi-scale resolution feature of CWT allows for the detection and quantification of the amplitude and timing (phase) of low-frequency streamflow discontinuities of pre-determined wavelengths. The Morlet wavelet, which is a Gaussian modulated sine wave, has been used as the mother wavelet for detecting discontinuities in streamflow due to its good basic frequency resolution (Massei *et al.*, 2011). Fraedrich *et al.* (1997) implemented a non-symmetric wavelet transform in a Mann-Kendall test and cluster-analysis based methodology to identify abrupt climate changes from flood levels of the Nile River.

Using XWT with the Morlet wavelet as the mother function has been shown to provide good time-frequency resolution suitable for water discharge or streamflow measurements (e.g. Labat, 2008, 2010). The stabilizing sinusoidal signal is incorporated by applying cross-wavelet analysis (XWT). Other types of wavelet transforms, such as the Mexican Hat and Haar wavelets, have also shown good local support (i.e. high accuracy in determination of temporal occurrences of cycles) in water discharge analysis (e.g. Smith *et al.* 1998) and may be applicable for discontinuity analysis.

This study focused on developing a methodology for detecting discontinuities in streamflow data using CWT and XWT, and tested the proposed method on a model data set and then applied it to streamflow data from Canadian rivers to identify patterns in streamflow discontinuities. In our proposed approach, we expand the approach of discontinuity detection of Jiang *et al.* (2002) to single hydrological or climate records by using, instead of a second record, an undisturbed stationary sinusoidal signal of the wavelength in which discontinuities are of interest. It should be noted that the Jiang *et al.* (2002) method is applied over all time scales to quantify the amplitudes and their significance. Our study is focused on specific time scales of interest and the counting of discontinuities that are above a specified threshold. Such temporal occurrences cannot be captured by student or other population, or autocorrelation-based tests. As such, our method does not include significance estimates comparable to those that were applied by Jiang *et al.* (2002).

## METHODOLOGY

The proposed method consists of the following four main steps:

1. Time series of waveband specific discontinuity magnitudes are extracted using the XWT of individual meteorological or streamflow records coupled with a sine wave. The wavelength of the sine wave is set to

the frequency of interest (e.g. 365 days, 1 day etc.), which is determined by the goals set for individual investigation, for example, the ‘impact of annual temperature variability on daily streamflow’.

2. Thresholds for discontinuity magnitudes are set from known abrupt record features such as artificial discontinuities arising from records gaps, or in the case of no gaps, above the mean discontinuity magnitude; above that, discontinuities are considered true and not noise related.
3. The number of discontinuities that are above the threshold are counted over a sliding time window and divided by the maximum number of possible discontinuities. This value is called the ‘discontinuity density’. This results in discontinuity density records in the rational number scale that can be evaluated with data analysis methods designed for continuous records.
4. Fourier analysis is used to determine periodic patterns and their significance in the discontinuity record, and these are then compared with the spectral power of periodicities in the underlying hydrological records.

The proposed XWT-based methodology was tested on the following: (i) a model data set that combines complex signals (described later in the paper); and (ii) six streamflow records. The testing included fine tuning of wavelet transform specific parameters and comparison to discontinuity detection limits of the CWT. For testing purposes, the methodology was compared with a CWT-only discontinuity extraction method proposed by Labat *et al.* (2000).

The proposed method can be used for a wide variety of discontinuities. Some examples include the following: records of natural processes, such as gradual streamflow fluctuations, are commonly interrupted by long or short disruptions from natural non-linear responses to gradual changes, for example, from river-ice break-ups, snow melt, flash floods from lake outbursts, freezing as a result of annual solar cycles or human causes such as dam opening/closing, spill overflows, saturation of wetlands closing of tributaries, or instrument calibrations and failure. The resulting abrupt or gradual shifts and missing data are considered to be discontinuities with respect to the normal signal.

## DATA ANALYSIS BACKGROUND

The method developed in this study for the detection and extraction of discontinuities in daily streamflow data involved discontinuity extraction based on the XWT for selected wavebands, which was found to be superior to extraction of discontinuities based solely on the CWT.

### Continuous wavelet analysis

The wavelet coefficients  $W$  of a time series  $x(s)$  are calculated by a simple equation

$$W_{\psi}(a, b) = \left(\frac{1}{\sqrt{a}}\right) \int x(s) \psi\left(\frac{s-b}{a}\right) ds \quad (1)$$

Where  $\psi$  is the mother wavelet,  $a$  is the scale factor that determines the characteristic frequency or wavelength and  $b$  represents the shift of the wavelet over  $x(s)$  (Prokoph and Barthelmes, 1996). The bandwidth resolution for a wavelet transform varies with  $\Delta a = \frac{\sqrt{2}}{4\pi a l}$  and a location resolution  $\Delta b = \frac{a l}{\sqrt{2}}$ . Because of Heisenberg’s uncertainty principle  $\Delta a \Delta b \geq 1/4\pi$ , the resolution of  $\Delta b$  and  $\Delta a$  cannot be arbitrarily small (e.g. Prokoph and Barthelmes, 1996). Parameter  $l$  is used to modify the wavelet transform bandwidth resolution either in favour of time or in favour of frequency.

In this study, the Morlet wavelet was used as the mother function (Morlet *et al.*, 1982), which is expressed in its shifted and scaled version as

$$\Psi_{a,b}^l(s) = \sqrt[4]{\pi} \sqrt{a l} e^{-\frac{i 2\pi(s-b)}{a}} e^{-\frac{1}{2}\left(\frac{s-b}{a l}\right)^2} \quad (2)$$

The Morlet wavelet is a sinusoid with wavelength/scale  $a$  modulated by a Gaussian function (Torrence and Compo, 1998; Adamowski *et al.*, 2009). Edge effects of the wavelet coefficients occur at the beginning and end of the analysed time series and increase with increasing wavelength (scale) and parameter  $l$ . Thus, a ‘cone of influence of edge effects’ is formed (Torrence and Compo, 1998). Influences of greater than 10% based on the wavelet analysis parameters used are illustrated in scalograms that are two-dimensional graphic presentations of the matrix of the wavelet coefficients  $W(a, b)$ , in the scale (or wavelength) time domain. The scalogram is coded with a colour scale (orange highest, blue lowest  $W(a, b)$ ) for better graphical interpretation (Figure 3).

The wavelet coefficients  $W$  are normalized using the L1 normalization ( $1/a$ ), replacing the commonly used  $1/\sqrt{a}$  for L2 or  $L^2$  normalization (see Equation 1). This allows wavelet coefficients to be interpreted in terms of Fourier amplitudes (e.g. Prokoph and Barthelmes, 1996). In addition, the L2 normalization of the Morlet wavelet commonly leads to overvaluing wavelet coefficients in long wavelengths compared with shorter ones (Schaeffli *et al.*, 2007).

Two parameters of  $l$  were tested for their precision in detecting wavelength-based discontinuities. A parameter of  $l=6$  provides sufficiently precise results in the resolution of time and scale and is commonly suggested for hydrological records (e.g. Schaeffli *et al.*, 2007;

Adamowski *et al.*, 2009), whereas  $l=2$  provides better time resolution but loses most of the frequency resolution (Rioul and Vetterli, 1991). In this study, the CWT was used to detect and extract the wavebands, magnitudes and phases of signals embedded in the hydrological records. Abrupt changes in these features indicate waveband-dependent discontinuities in single records; these were then compared with the results of discontinuities extracted using the XWT (see below).

*Cross-wavelet analysis (XWT)*

The cross-wavelet spectrum  $W_{xy}$  of two series  $x(s)$  and  $y(s)$  is defined by

$$W_{xy}(a, b) = W_x(a, b)W_y^*(a, b) \tag{3}$$

where  $W_x(a, b)$  and  $W_y(a, b)$  are the CWT of  $x(s)$  and  $y(s)$ , respectively, and  $*$  denotes the complex conjugate (e.g. Jury *et al.*, 2002). The phase difference is defined by

$$\Delta\Phi(b) = \tan^{-1} \frac{\int_{a1}^{a2} \text{Im}(W_{xy}(a, b)) da}{\int_{a1}^{a2} \text{Re}(W_{xy}(a, b)) da} \tag{4}$$

*Im* and *Re* indicate the imaginary and real parts, respectively. The mother wavelet and parameters used in this study are the same as for the CWT description provided above. More details on the XWT technique used in this study can be found in Maraun and Kurths (2004) and Labat (2008). As for the CWT in this study, the XWT applies the L1 normalization. Consequently,  $W_{xy}$  represents the cross amplitude between a stationary sinusoidal record with the wavelength of interest for discontinuity extraction and the real or modeled records.

*Discontinuity extraction*

A function  $f(s)$  has a discontinuity of degree  $k$  at  $s_0$ ,  $k=0, 1, \dots$ , if the  $k$ -th order left and right derivatives at  $t_0$  are different; that is,  $f^k(s_0+) \neq f^k(s_0-)$ , the difference being the size of the discontinuity (Lee 1989). Using the CWT (see Equation 1) according to this definition, a discontinuity record  $u$  for each  $b$  between the time intervals  $b-1$  and  $b+1$  in a single time series can be interpreted as

$$u(b) = \sqrt{\sum_{a=1}^n \left( \frac{W(a, b)}{\sum W(b)} - \frac{W(a, b-1)}{\sum W(b-1)} \right)^2} \tag{5}$$

where  $n$  is the total number of frequencies. For waveband-specific discontinuity detection, the range of  $a=1$  to  $N$  can

be narrowed to a waveband of interest. The software used in this study (developed by Prokoph and Barthelmes, 1996) allows one to define the waveband of interest.

Extending the equation for discontinuity definition from one to two time series and using notations for XWT (Equation 3), the discontinuities  $v$  between two time series  $x$  and  $y$  can be accordingly expressed as

$$v(b) = \sqrt{\sum_{a=1}^n \left( \frac{W_{xy}(a, b)}{\sum W_{xy}(b)} - \frac{W_{xy}(a, b-1)}{\sum W_{xy}(b-1)} \right)^2} \tag{6}$$

The waveband of interest can be set by the user in the same way as for the CWT (see above). This waveband-specific XWT permits one to extract discontinuity records that are unaffected by high-frequency ('white') noise or long-term trends ('red noise') (Figure 3).

*Model data set*

A model data set was developed with the purpose of evaluating the accuracy of the waveband restricted discontinuity detection in a controlled environment with well-defined noise and gradual and abrupt short-term and long-term changes that occur at different wavelengths. The model data set has the following features (Figure 2A):

1. A stationary cosine background signal of annual wavelength (i.e. 365.25 days), which reflects a common annual streamflow cycle in Canadian rivers (Environment Canada, 1999);
2. A random high-frequency (white) noise at five times the amplitude of the periodic background signal (the resulting 1/5 signal-to-noise ratio limits the opportunity to detect the periodic signal by visual inspection);



Figure 1. Location of streamflow stations used with RHBN code (Table I)



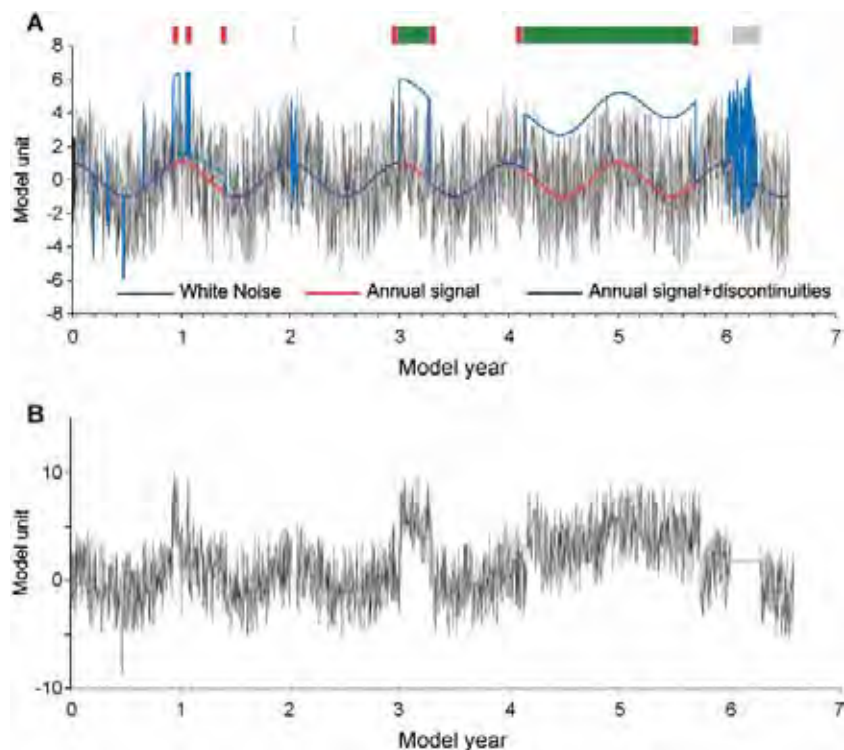


Figure 2. Model data set of A, constituting four features (overlying 1-year cosine, white noise, spikes and level shifts), with the red bar indicating the onset and offset of discontinuities and the green bars showing the range of offset. B: Complete model including additions of gaps record (for details, see text)

3. Gaps in measurements of varying durations (see Figures 2B) (such gaps reflect missing/dismissed measurements mostly because of recording errors or suspension caused by administrative issues);
4. Short-term spikes and abrupt shifts  $>$  magnitude 1 of the mean values compared with the 100 days mean value; and
5. Abrupt shifts over 1 day and gradual shifts over 30 days or more with a subsequent return to normal conditions with magnitude 1, which simulates the influence of dams or changes in the drainage system.

The duration ( $\sim 30$ – $100$  days) of the modeled shifts are based on the goal of finding patterns that are not confused with high-frequency noise of no long-term consequence or with long-term (multi-seasonal) shifts that will form gradual changes (trends) not subject to response to discontinuities but to long-term adaptation/changes. The magnitude of 1 in the presence of a noise magnitude of 5 was chosen to avoid detection of discontinuities by simple observation. The complete model data set, which incorporates the above described five features, is shown in Figure 2B.

The CWT of a simple cosine signal results in better wavelength (frequency) resolution using an analysis window size of  $l=6$ , whereas  $l=2$  has smaller edge effects (Figures 3A and 3B), in accordance with the findings of previous research (e.g. Rioul and Vetterli,

1991). The testing of different wavelet analysis window sizes is essential to determine the optimal scale-to-time resolution because of Heisenberg's uncertainty (see above). The CWT of the more complex full model data set with the same choice of  $l$  illustrates that, despite the shortcomings in wavelength resolution, the use of  $l=2$  enables better detection of discontinuities in the record (Figures 3C and D). In particular, the abrupt, large shift towards higher model values at 4.1 model years (Figures 2A and 3K) is well represented by an abrupt change in wavelet coefficients and wavelength with  $l=2$  (Figures 3C) but not with  $l=6$  (Figure 3D).

Many discontinuities are overprinted by white noise, particularly in short wavelengths. The XWT with the stationary annual cosine wave and the model enhances the discontinuities in the annual wave band for  $l=2$  but smoothes out most discontinuities when  $l=6$  is applied (Figures 3E, 3F and 3I). The XWT discontinuities extracted from the [0.5 year, 2 year] wavebands (Figures 3G and 3H) indicate that for  $l=6$ , only a single broad positive discontinuity anomaly occurs at the end of the model record covering the multi-year model-data gap (Figures 3J). In contrast, nine strong, well-localized ( $<0.5$  years peak width) discontinuity anomalies are detectable using  $l=2$  that can be mostly related to discontinuities of greater than 0.5-year duration (Figures 3J and 3K). The anomaly at each end of the model record is unrelated to the discontinuities and is



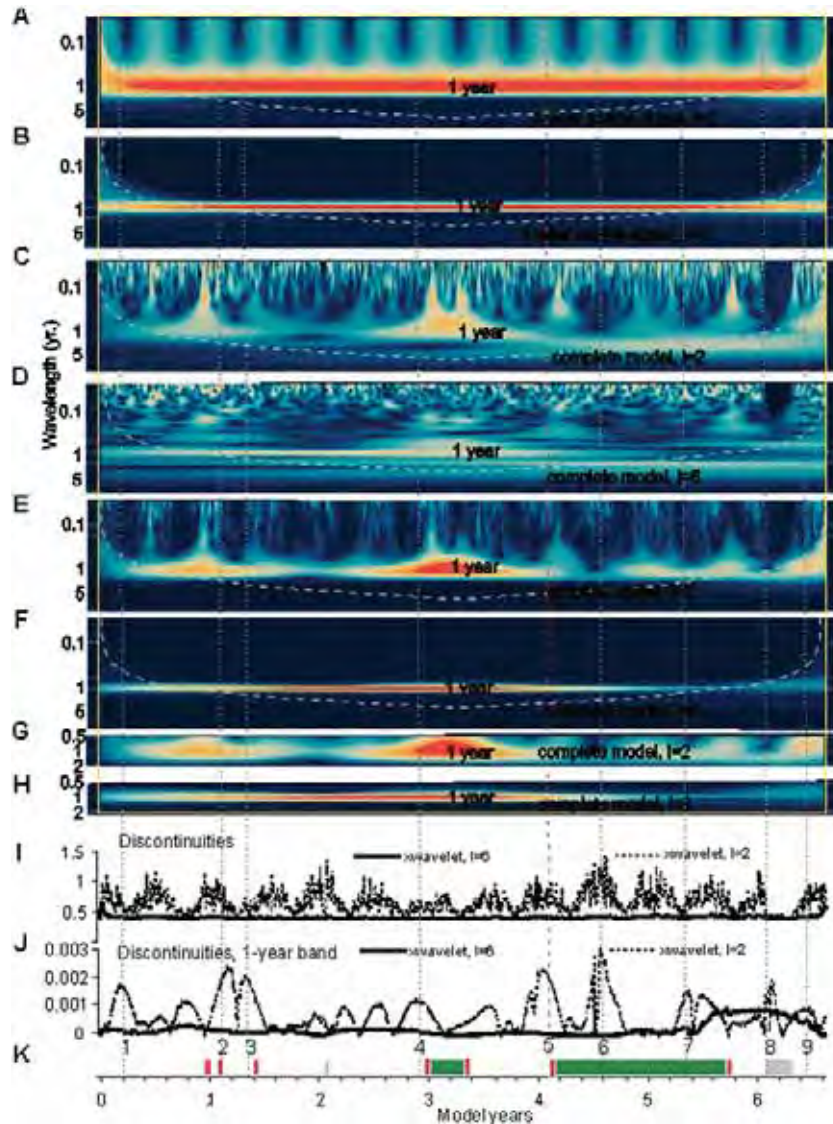


Figure 3. Steps of discontinuity detection from model data set (Figure 2B). Dashed lines show cone of influence (see text); vertical dashed lines delineate major discontinuities. Scalograms of A, CWT of cosine component using  $l=2$ ; B, CWT of cosine component using  $l=6$ ; C, CWT of complete model using  $l=2$ ; D, CWT of complete model using  $l=6$ ; E, XWT of complete model using  $l=2$ ; F, XWT of complete model using  $l=6$ ; G, XWT of complete model for [0.5,2] year wavebands using  $l=2$ ; H, XWT of complete model for [0.5,2] year wavebands using  $l=6$ ; I, XWT-discontinuities for entire waveband; and J, XWT-discontinuities for [0.5,2] year waveband

likely to be related to edge effects. Consequently, discontinuities can be best detected with a narrow analysis window and in record parts that are not influenced by edge effects.

### STUDY SITES AND DATA

The proposed method was first tested on a simple cosine signal and then on the model data set to compare the precision of the different values of parameter  $l$  in detecting wavelength-based discontinuities. The method was then tested on Canadian daily streamflow records from different geographic locations. The hydrological

data used in this study were extracted from the Reference Hydrometric Basin Network (RHBN) data sets produced by Environment Canada (1999), which include over a hundred records of annual maximum flows in Canadian rivers. The regularly extended sunspot number record ([www.noaa.gov](http://www.noaa.gov)) was used for comparison with hydrological signals and discontinuities.

Six of the watersheds belonging to the RHBN were selected, from which daily streamflow data from a specific gauging station on each river were used (Table I and Figure 1). The watersheds were selected on the basis that they are pristine (i.e. unaltered by dams, canals or agricultural activity) and that data exist

Table I. Locations, ecological settings and record ranges of hydrological stations

ID RHCBN	ID EC*	Location	Drainage in km <sup>3</sup>	Region	Time range A.D.
85	01AD002	Saint John River at Fort Kent	14 700	Atlantic Maritime	1926–2009
165	08GA010	Capilano River above Intake	172	Pacific Maritime	1914–1955,1958–2009
29	05PB014	Turtle River near Mine Centre	4870	Boreal Shield	1920–1979,1984–2009
4	05AD005	Belly River near Mountain View	319	Montane Cord.	1912–2010
6	05BB001	Bow River at Banff	2210	Montane Cord.	1910–2010
X1	02KF006	Mississippi River at Appleton	2900	Mixedwood Plain	1918–2009

\*EC Environment Canada

for over 40 years. It was also ensured that five of the southern Canadian ecozones were represented (Table I). The ecozones are defined based on vegetation, wildlife, latitude, altitude, proximity to oceans, terrain and other criteria (Environment Canada, 1999). However, the Mississippi River at Appleton (watershed X1) is not part of the RHBN but was included in the selection as it is the only long-term daily record representative of its ecozone.

Because the watersheds were chosen to represent five different ecozones, there is considerable variation between them in terms of climate and patterns of precipitation and streamflow. Eastern Ontario and southern Quebec (watersheds 85 and X1) have a continental climate with long, cold winters and short, humid summers (Zhang *et al.*, 2001; Adamowski *et al.*, 2012). The streamflow pattern in these regions is generally charac-

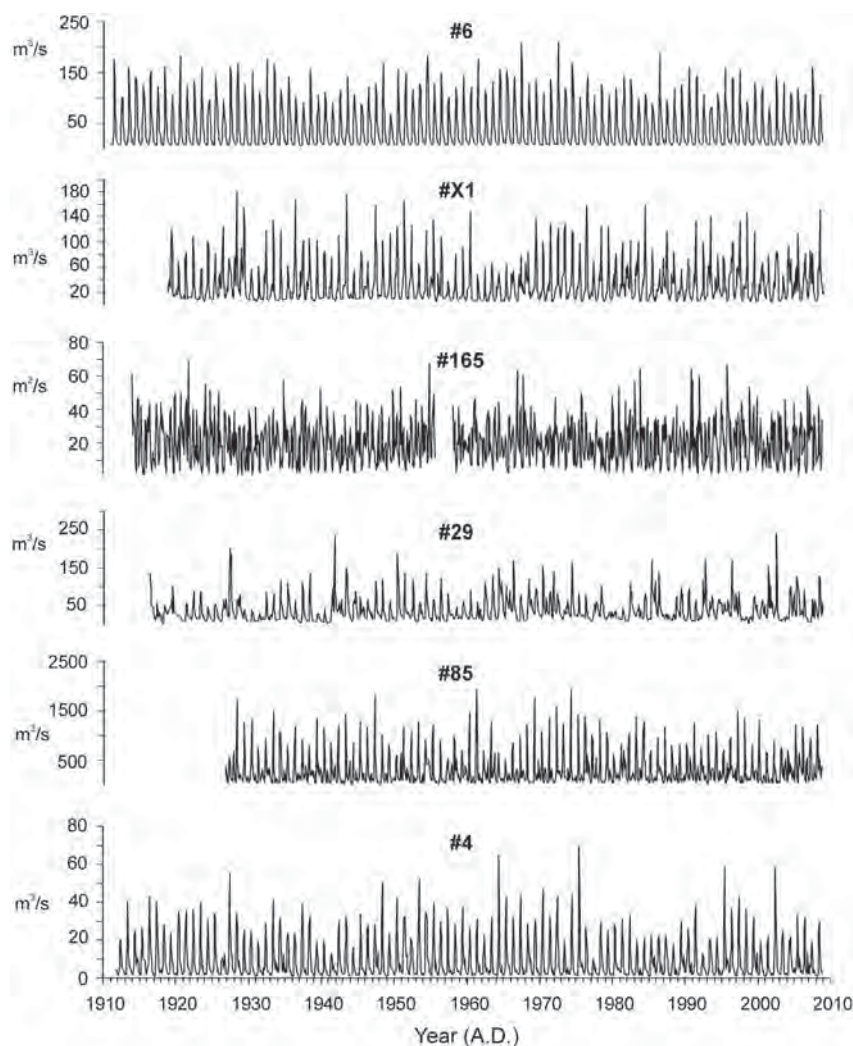


Figure 4. Canadian daily streamflow records. Station locations are shown in Table I and Figure 1

terized by a spring flood related to snowmelt and relatively low flows during the remainder of the year (Adamowski, 2008). The Canadian Shield area of north-western Ontario (watershed 29) has a strongly continental climate, with large seasonal differences in temperature. The climate is relatively moist, with the summer months (i.e. between May and September) being the wettest. The majority of precipitation outside this period falls as snow (St George, 2007). By contrast, southern Alberta (watershed 4) experiences a cool, semi-arid climate, as it lies in the rain shadow of the Rocky Mountains (Byrne *et al.*, 1989). Watershed 6 at Banff, Alberta experiences a similar climate as it is located on the eastern side of the Rocky Mountains.

Finally, coastal British Columbia (watershed 165) has a more maritime climate with moderate temperatures and high precipitation that predominantly falls as rain rather than snow during the winter (Shaefer, 1978).

RESULTS AND DISCUSSION

Statistical tests based on the recommendations of Shiao and Condie (1980) were carried out to verify the quality of the RHBN data (Environment Canada, 1999). All of the records start before 1930 and last until at least 2009, with two of the records exhibiting several gaps (Figure 4,

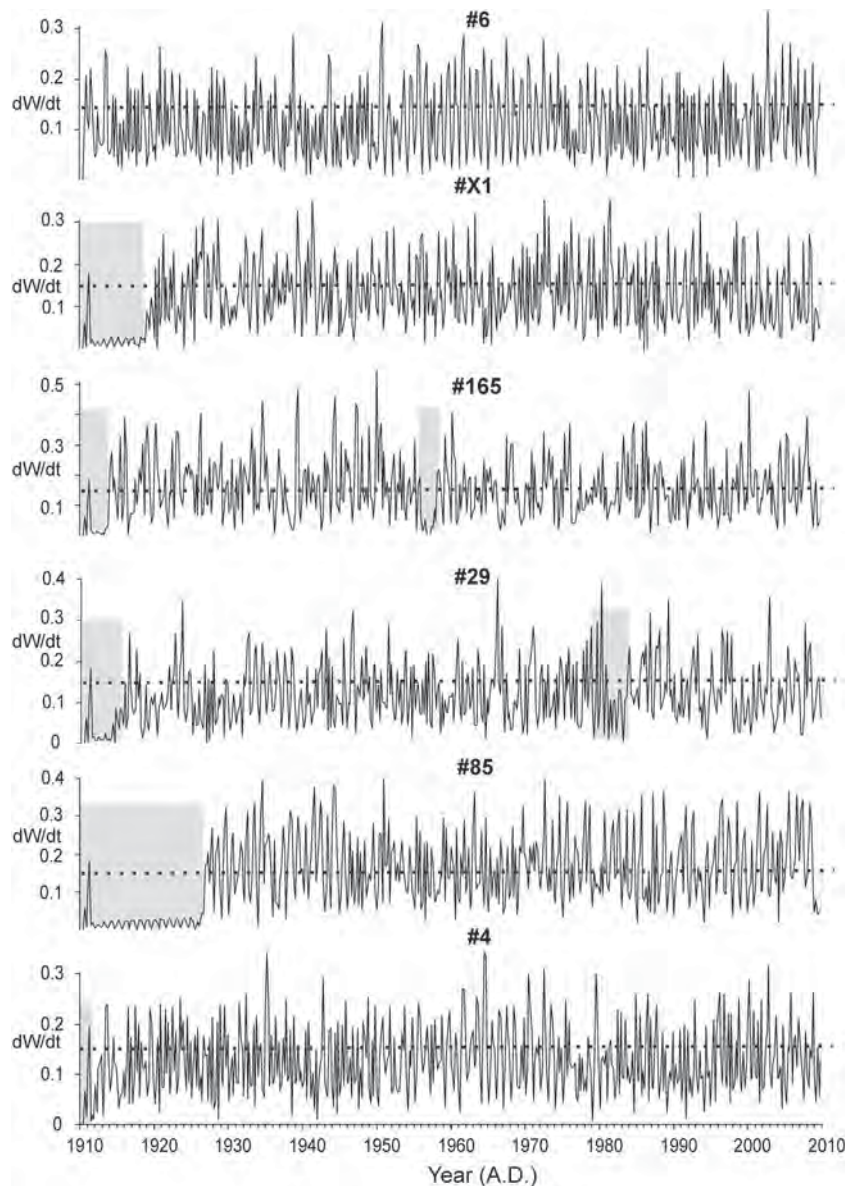


Figure 5. Discontinuities ( $g = dW/dt$ ) of daily streamflow data in  $[0.5, 2]$  year waveband using  $l=2$ . Gray shaded areas mark missing data in the records. Dashed line delineates discontinuity threshold of  $dW/dt=0.15$  that has been used for discontinuity density calculations



Table I). Records of sunspot data corresponding to the same time interval as the streamflow records were collected from the NOAA Web page.

At first, the annual waveband ([0.5, 2] year) was investigated for discontinuities because fluctuations on an annual scale exhibit the most important fluctuations in Canadian streamflows, such as spring flood due to snow melt and summer droughts (e.g. Zhang *et al.*, 2001). Discontinuity analysis for the [0.5, 2] year waveband shows that missing data at the beginning and in the middle of the records are easily identified by low discontinuity values (Figure 5). As expressed in the

methodology, the discontinuity threshold is best obtained from  $dW/dt$  across data gaps. Based on this definition, a threshold discontinuity value of 0.15 in the annual waveband was derived for all records based on data gaps and mean discontinuity value. The mean discontinuity value ranges from 0.11 to 0.15 despite the different hydrological settings for the streamflows. The similar range of the discontinuity from 0 to  $\sim 3.5$   $dW/dT$  (Figure 5) also indicates that the same discontinuity threshold for all analysed record is reasonable. Only record times that have discontinuity values above 0.15 had a nominal discontinuity count assigned (yes = 1, no = 0).

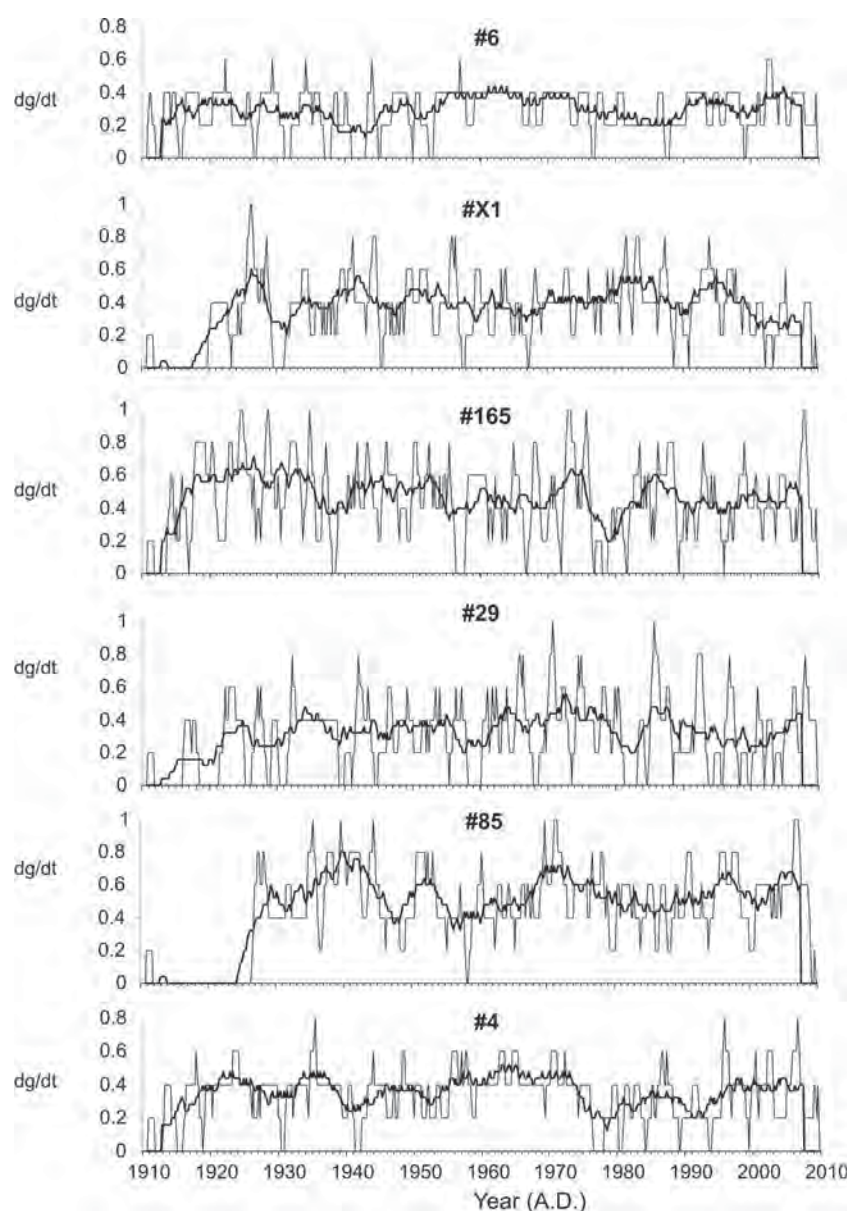


Figure 6. Discontinuity density ( $dg/dt$ ) for all records averaged over 1 and 5 years. Note that the parts of sections that have data missing were excluded from averaging

A single large discontinuity was observed at the beginning of the record for five of the six watersheds, indicating an edge effect response (Figure 4J). Otherwise, no obvious pattern in discontinuity is discernible, apart from correspondence between watersheds 4 and 6, which are located within the same ecozone (Table I and Figure 1). These watersheds exhibited only one discontinuity above 0.15 per year in the 1960s (Figure 5).

The nominal discontinuity assignments above the threshold of 0.15 were calculated over 1- and 5-year windows, providing a maximum resolution for the 1-year waveband and a high-robustness estimate, respectively. The discontinuity densities were high in the late 1950s and 1970s, and low in the late 1970s in all records, and generally show a pronounced cyclical nature (Figure 6). The 5-year average of all discontinuity density records exhibits a pronounced 11-year periodicity. Cross-correlation with the sunspot number record shows that the discontinuity density in Canadian streamflows is approximately inversely correlated with the ~11-year periodic solar intensity variability (Figure 7). The total averaged streamflow from the selected Canadian stations generally also decreases when there are less discontinuities, although the decrease is less obvious (Figure 7). In addition, an 11-year cyclicity in total averaged streamflow is subdued by a stronger bi-decadal cyclicity (Figures 7 and 8). In this particular example, the results indicate that more discontinuities occur during years of decreased solar radiation and that there is a weak correlation between the densities of discontinuities and the streamflow in general. We argue that during lower solar intensity years, the seasonal climate regimes change more abruptly in mid-latitude settings in Canada. The subsequent more rapid onsets of freezing and ice break-up, as well as more

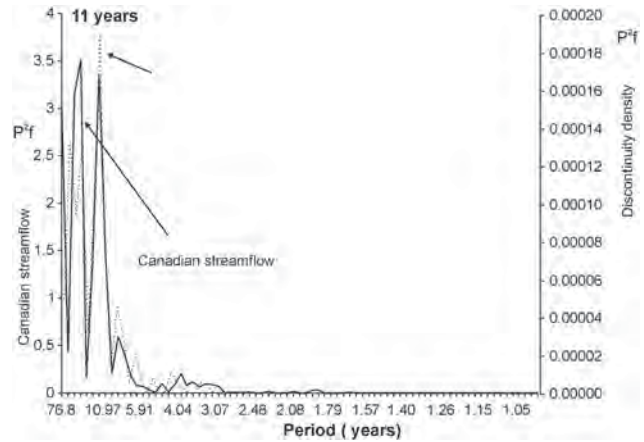


Figure 8. Spectral analysis of 5-year averaged streamflow record from the selected Canadian stations (blue) and discontinuity density records (red)

frequent short-term weather fluctuations, may lead to more abrupt changes in the streamflow pattern as shown with larger spring floods in the Rocky Mountains of Canada during low solar intensity years (Prokoph *et al.*, 2012).

The proposed methodology of discontinuity detection and analysis could also be used to evaluate the effects of other natural (e.g. El Nino, monsoon, streamflow change because of landslides) or human-made (e.g. waterway widening and cleaning) phenomena on abrupt changes in streamflow patterns. Furthermore, the proposed method allows one to remove the influence of already known discontinuities (e.g. known streamflow magnitude jumps due to dam opening and closing). This allows one to focus on detection and interpretation of discontinuities occurring in the waveband of interest.

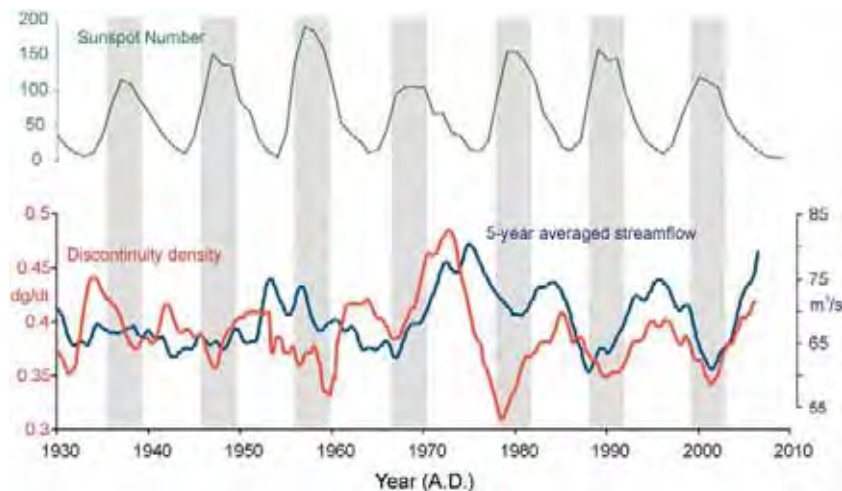


Figure 7. Annual sunspot numbers (source: www.noaa.gov), 5 years averaged discontinuity density of all streamflow records, and 5 years averaged streamflow record from the selected Canadian stations



## CONCLUSION

The proposed methodology is designed to detect streamflow discontinuities that are significant for periods (wavelengths) of interest. Practical examples of such discontinuities would be those due to dam opening/closing, spill overflows, saturation of wetlands or closing of tributaries. It is shown that the XWT of the observed record as the first time series and a stationary sinusoidal signal of wavelength of interest can detect and extract such discontinuities specific for the wavelength of interest. In particular, discontinuities at the onset and offset of a shift in the mean of the records for at least half of the duration of the wavelength of interest can be detected and separated from other signals and noise. Parameter testing of the time-frequency resolution of the Morlet wavelet shows that high time resolution is favorable to high-frequency resolution for detection of waveband-related discontinuities.

The application of the proposed methodology on a model data set shows that the strength and timing of abrupt shifts to a new lower or higher streamflow level, gaps longer than the waveband of interest, and a sinusoidal discontinuity curve following an underlying modeled annual signal are detected at  $\pm 0.5$  year uncertainty.

Discontinuity analysis on observed daily streamflow records shows the following: (i) that each record has at least one discontinuity/year related to the annual spring-flood, (ii) that neighboring streamflows have similar discontinuity patterns, and (iii) the discontinuity densities in Canadian rivers follow an 11-year cycle inverse to the solar intensity cycle. An 11-year cycle is also evident in the raw streamflow records but is less pronounced.

The proposed methodology was applied to streamflows in Canada but can be similarly used in other parts of the world. This could include the effects of the onset and end of the monsoon or changes in the onset of annual snowmelt over time. Other wavelet types such as the Haar wavelet may also be useful for this type of detection because of their excellent local support. This study focused primarily on discontinuities in the annual streamflow pattern but could be further used for studies to evaluate the effect of discontinuities on other time scales, for example, the influence of abrupt changes of oceanic-atmospheric oscillation patterns on forming abrupt streamflow pattern changes. The methodology that is proposed and presented in this paper is able to extract streamflow discontinuities in specified wavelength and defined thresholds. For future studies, a method of testing the significance of the thresholds would be beneficial.

## ACKNOWLEDGEMENT

Financial support for this study was provided by a Natural Sciences and Engineering Research Council of Canada (NSERC) and a CFI grant held by Jan Adamowski.

## REFERENCES

- Adamowski J. 2008. Development of a short-term river flood forecasting model for snowmelt driven floods based on wavelet and cross-wavelet analysis. *Journal of Hydrology* **353**: 247–266.
- Adamowski K, Prokoph A, Adamowski J. 2009. Development of a new method of wavelet aided trend detection and estimation. *Hydrological Processes* **23**: 2686–2696.
- Adamowski JF, Prokoph A, Adamowski K. 2012. Spatial temporal changes in streamflow patterns in eastern Ontario and southwestern Quebec, Canada and their relation to precipitation changes. *International Journal of Climate Change: Impacts and Responses* **3**(1): 155–170.
- Byrne JM, Barendregt R, Schaffer D. 1989. Assessing potential climate change impacts on water supply and demand in southern Alberta. *Canadian Water Resources Journal* **14**(4): 5–15.
- Environment Canada. 1999. Establishment of the reference hydrometric basin network (RHBN) for Canada. *Environment Canada, Ottawa*: 42 p.
- Fraedrich K, Jiang J, Gerstengarbe W. 1997. Multiscale detection of abrupt climate changes: Application to Nile River flood levels. *International Journal of Climatology* **17**: 1301–1315.
- Goossens C, Berger A. 1987. How to recognize an abrupt climatic change?, In *Abrupt Climatic Change*, Berger WH, Labeyrie LD (eds). D Reidel Publisher: Dordrecht 31–45.
- Jiang J, Mendelsohn R, Schwing F, Fraedrich K. 2002. Coherency detection of multiple abrupt changes in historic Nile flood levels. *Geophysical Research Letters* **29**: 1121–1124.
- Jury MR, Enfield DB, Mélice J. 2002. Tropical monsoons around Africa: Stability of El Niño–Southern Oscillation associations and links with continental climate. *Journal of Geophysical Research* **107**: 3151–3167.
- Kunhel I, McMahon TA, Finlayson BL. 1990. Climatic influences on streamflow variability: A comparison between southeastern Australia and southeastern United States of America. *Water Resources Research* **26**: 2483–2496.
- Labat D. 2008. Wavelet analysis of the annual discharge records of the world's largest rivers. *Advances in Water Resources* **31**: 109–117.
- Labat D. 2010. Cross wavelet analyses of annual continental freshwater discharge and selected climate indices. *Journal of Hydrology* **385**: 269–278.
- Labat D, Ababou R, Mangin A. 2000. Rainfall-runoff relations for karstic springs. Part II: Continuous wavelet and discrete orthogonal multiresolution analyses. *Journal of Hydrology* **238**: 149–178.
- Lee L, Huang L, Chao J. 1989. On the stability of rotational discontinuities and intermediate shocks. *Journal of Geophysical Research: Space Physics* **94**: 8813–8825.
- Mallat S, Hwang WL. 1992. Singularity Detection and Processing with Wavelets. *IEEE Transactions on Information Theory* **38**: 617–639.
- Maraun D, Kurths J. 2004. Cross wavelet analysis: significance testing and pitfalls. *Nonlinear Processes in Geophysics* **11**: 505–514.
- Massei N, Laignel B, Rosero E. 2011. A wavelet approach to the short-term to pluri-decadennial variability of streamflow in the Mississippi river basin from 1934 to 1998. *International Journal of Climatology* **31**: 31–43.
- Morlet J, Arehs G, Fourgeau I. 1982. Wave propagation and sampling theory. *Geophysics* **47**: 203–236.
- Nguyen DT, Hoang TA. 1999. Detection of disturbances on electricity supply using wavelets. *Australasian Universities Power Engineering Conference and IEAust Electric Energy Conference*: 184–189.
- Pettitt AN. 1979. A nonparametric approach to the changepoint problem. *Applied Statistician* **28**: 126–35.
- Prokoph A, Barthelme F. 1996. Detection of nonstationarities in geological time series: Wavelet transform of chaotic and cyclic sequences. *Computer & Geosciences* **22**: 1097–1108.
- Prokoph A, Adamowski J, Adamowski K. 2012. Influence of the 11 year solar cycle on annual streamflow maxima in Southern Canada. *Journal of Hydrology* **442–443**: 55–62.
- Rioul O, Vetterli M. 1991. Wavelets and Signal Processing. *IEEE Signal Processing Magazine* **8**: 14–38.
- Schaeffli B, Maraun D, Holschneider M. 2007. What drives high flow events in the Swiss Alps? Recent developments in wavelet spectral analysis and their application to hydrology. *Advances in Water Resources* **30**: 2511–2525.
- Shafer DG. 1978. Climate. In *The soil landscapes of British Columbia*, B. C., Valentine KWG, Sprout PN, Baker TE, Lavkulich LM (eds). The

- soil landscapes of British Columbia, B.C. Ministry of Environment, Resource Analysis Branch: Victoria; 197.
- Shao Q, Li Z, Xu Z. 2010. Trend detection in hydrological time series by segment regression with application to Shiyang River Basin. *Stochastic Environmental Research and Risk Assessment* **24**: 221–233.
- Shiao SY, Condie R. 1980. Statistical tests for independence, trend, homogeneity and randomness. Inland Waters Directorate, Environment Canada.
- Smith LC, Turcotte DL, Isacks BL. 1998. Stream flow characterization and feature detection using a discrete wavelet transform. *Hydrological Processes* **12**: 233–249.
- St George S. 2007. Streamflow in the Winnipeg River basin, Canada: Trends, extremes and climate linkages. *Journal of Hydrology* **332**: 396–411.
- Torrence C, Compo GP. 1998. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society* **79**: 61–78.
- Zhang X, Harvey K, Hogg WD, Yuzyk R. 2001. Trends in Canadian streamflow. *Water Resources Research* **37**: 987–998.
- Zhao F, Xu Z, Zhang L. 2011. Changes in streamflow regime following vegetation changes from paired catchments. *Hydrological Processes* Doi:10.1002/hyp.8266

---

## First Nation capacity in Québec to practise integrated water resources management

---

Zehra Rizvi and Jan Adamowski\*

Department of Bioresource Engineering, McGill University,  
2111 Lakeshore, Ste Anne de Bellevue, Quebec, H9X3V9, Canada  
E-mail: zehra.rizvi@mail.mcgill.ca  
E-mail: jan.adamowski@mcgill.ca  
\*Corresponding author

Robert J. Patrick

Department of Geography and Planning,  
University of Saskatchewan, Saskatoon, SK, S7N 5C8, Canada  
E-mail: robert.patrick@usask.ca

**Abstract:** Integrated Water Resources Management (IWRM) has been identified by the United Nations as a critical component of effective and sustainable water resources management in the future. This research examined the extent to which IWRM is practised among First Nations (FN) in Canada. This study also developed and applied an analytical framework to assess the overall capacity of two FN communities in Québec to practise IWRM. The FN communities of Kitigan Zibi and Kahnawà:ke were evaluated with respect to capacity to support actor network, information management, human resources and technical, financial and institutional dimensions. This study recommends that future Québec IWRM initiatives with FN collaboration be directed towards strengthening actor network capacities and understanding the complexity of FN perspectives. In addition, the results of this study indicate that FNs with limited financial capacity will experience reduced actor network, information management, human resources and technical capacity.

**Keywords:** First nations; indigenous; capacity; IWRM; integrated water resources management; Québec; Canada.

**Reference** to this paper should be made as follows: Rizvi, Z., Adamowski, J. and Patrick, R.J. (2013) 'First Nation capacity in Québec to practise integrated water resources management', *Int. J. Water*, Vol. 7, No. 3, pp.161–190.

**Biographical notes:** Zehra Rizvi holds a MSc in Integrated Water Resources Management (IWRM) from the Department of Bioresource Engineering at McGill University in Canada. Her applied and research interests include: participatory processes of watershed and local-level water management; collaborative stakeholder partnerships; socio-economic impact assessments at the local level; water, sanitation and hygiene (WATSAN) tools and techniques; and the design and application of integrated analytical frameworks. She has interest in working with indigenous communities, women and the Global South. Prior to her graduate research, She obtained her BA (Honours) in International Studies from the University of Saskatchewan.

Jan Adamowski is an Assistant Professor in hydrology and water resources management in the Department of Bioresource Engineering at McGill University in Canada. He is also the Director of the Integrated Water Resources Management Program and the Associate Director of the Brace Centre for Water Resources Management at McGill. His research interests include: participatory and integrated modelling and assessment of watershed policies and strategies through the use of system dynamics modelling; modelling and forecasting of nonlinear and non-stationary hydro-meteorological time series; and hydro-meteorological trend detection and estimation. Prior to coming to McGill, he was a Post-doctoral Associate at MIT in the USA.

Robert Patrick is Chair of the Regional and Urban Planning Program in the Department of Geography and Planning in the College of Arts and Science at the University of Saskatchewan in Saskatoon, Canada. His research area of interest is watershed planning for source water protection. He undertakes research in Saskatchewan and the Northwest Territories with both government and FNs in Canada. He has developed a local community engagement model for source water protection planning. He teaches watershed planning and management and regional planning courses at the University of Saskatchewan.

---

## 1 Introduction

A majority of the world's Indigenous peoples within nation-states are rarely involved as collaborators in meaningful discussions of water policies (United Nations, 2009). A correlation has been suggested to exist between ethnic indigenous identity and limited access to water (Macisaac, 1996; Gracey et al., 1997; Bailie et al., 2004; United Nations, 2009). Integrated Water Resources Management (IWRM), a paradigm shift in the management of water resources, could potentially reduce water inequities between users and increase Indigenous participation. IWRM embraces principles of stakeholder participation in decision-making, equity of water allocation, efficient and balanced water use and recognition of linkages and interactions among human and physical systems (Global Water Partnership, 2000). Canada is in the process of evolving from traditional approaches to water management, which can be characterised as fragmented, engineering-based, supply-oriented, sectoral, reactive and top-down, to more integrated and collaborative approaches (Ramin, 2004; Mitchell, 2006).

Canada's progress in IWRM is expanding: several provinces, including Québec, have developed comprehensive provincial water policies. Stakeholder participation is often seen as an essential vehicle to achieve short and long-term goals in integrated and collaborative water management (Dalton, 2006; Watson, 2007; Morin and Cantin, 2009; Roy et al., 2009). In 2002, Québec developed a comprehensive strategy to formalise IWRM and stakeholder participation (MDDEP, 2002a): the Québec Water Policy (QWP). This strategy began by identifying 33 watershed organisations (ROBVQ, 2010), whose mandates were to develop a watershed management plan and act as regional round tables to which any and all water resource stakeholders were invited (ROBVQ, 2010). In March 2009, the QWP expanded to cover 40 watershed zones with a focus on southern Québec (MDDEP, 2002a, ROBVQ, 2010).<sup>1</sup> Given the QWP's recognition of water as a "valuable asset of Québec society and an integral part of its collective heritage," water governance reforms include strengthening Québec's partnerships and ensuring all water

management players, particularly aboriginals, are involved in achieving the province's water management goals (MDDEP, 2002b).<sup>2</sup>

While the importance of capacity is widely acknowledged in IWRM, more attention needs to be drawn to holistic evaluations of indigenous communities' ability to practise IWRM. The existing literature mainly evaluates capacity for implementation of specific aspects of IWRM, including:

- source water protection or management (De Loë et al., 2002; De Loë and Lukovich, 2004; Carter et al., 2005; Ivey et al., 2006; Timmer et al., 2007)
- desalination (Al-Jayyousi, 2000)
- IWRM implementation at the national level (Mkandawire and Mulwafu, 2006)
- urban water management (Brown, 2008)
- drought (Hundertmark, 2008)
- rain water harvesting (Farahbakhsh et al., 2009)
- institutional capacity (Lamoree and Harlin, 2002).

Although studies specific to the capacity of Canada's indigenous people to implement IWRM exist (Smith et al., 2006; Lebel and Reed, 2010), there remains insufficient discussion on the capacity of indigenous communities to holistically practise IWRM in Canada. To date, Lebel and Reed (2010) evaluated the capacity of a Saskatchewan First Nation (FN) community to provide safe drinking water in terms of financial and human resources and institutional, socio-political and technical aspects, while Smith et al. (2006) investigated 56 FN drinking water systems in Alberta with regard to technical and human resources. Indigenous indicators have yet to be created in the literature. It is hoped that future research will explore indigenous indicators in IWRM. At present, there is an abundance of literature that examines capacity as it relates to small water systems and which discusses financial, technical and other capacity areas.

The purpose of this study was to develop and apply an analytical framework and then report findings for Kitigan Zibi and Kahnawà:ke FNs to practise IWRM in the province of Québec, particularly in the areas of actor network, information management, human resources and technical, financial and institutional capacities.

## **2 Study areas**

Selection of participant communities was based on:

- geographical representation of rural and urban regions
- situated within the province of Québec
- agreement on study objectives and research processes from Band Chief or representatives
- having a FN designation as defined by the Government of Canada.

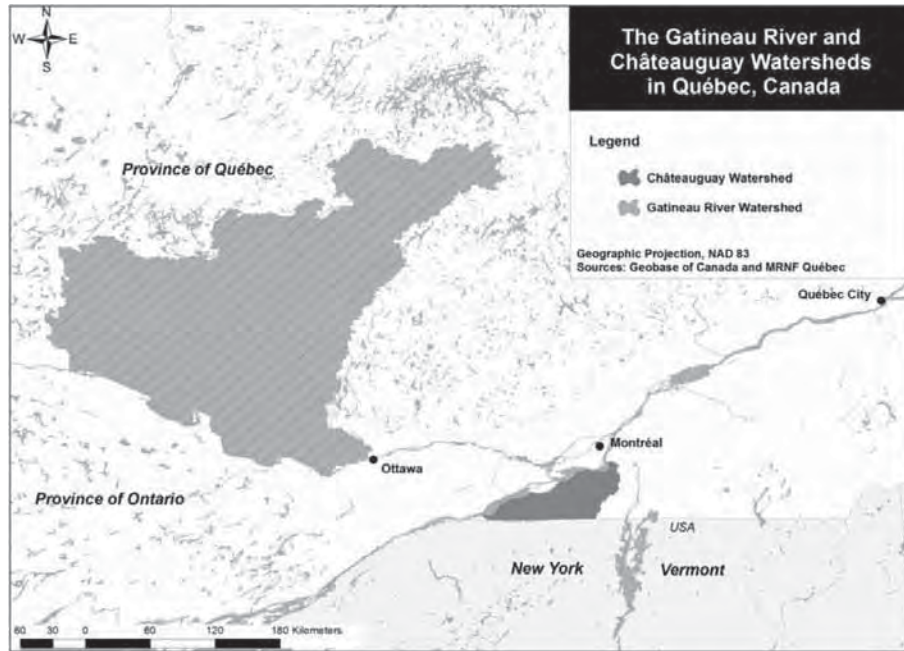
### *2.1 Kitigan Zibi FN*

Kitigan Zibi is a rural community of 1,557 Algonquin residents, located 130 km north of Gatineau, Québec. It is bound on the north by Rivière de l'Aigle and Rivière Désert

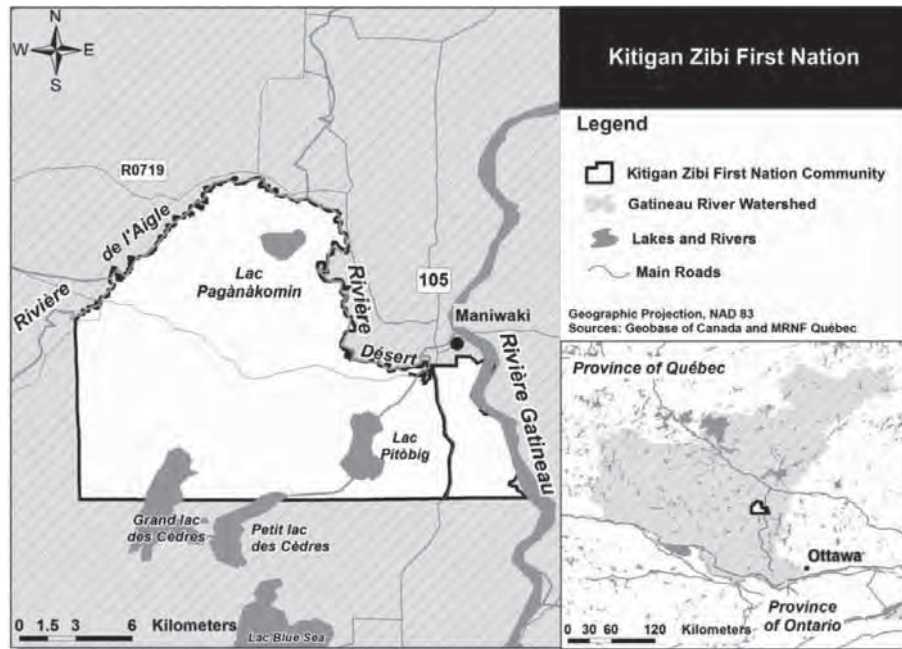


(Figure 1). The community covers 18,438 hectares (45,559 acres) and is part of the Gatineau River Watershed (Figures 1 and 2). Based on the geographical 'remoteness factor', there is year-round road access and it is located less than 50 km from the nearest service centre (INAC, 2010b). There are about 25 small businesses in Kitigan Zibi (INAC, 2010b).

**Figure 1** The Gatineau river and Châteauguay watersheds in Quebec, Canada



At the time of the study, some 88 Kitigan Zibi households (or 17% of total households) were connected to the piped water distribution and wastewater system of Maniwaki, a neighbouring non-aboriginal municipality. Maniwaki's piped water distribution relies on groundwater. The 437 remaining homes (or 83% of total households) used well-distribution systems for non-drinking purposes, relied on INAC-supplied bottled water for drinking and had their own sewage fields (Kitigan Zibi FN, personal communication, 30 April, 2010). Groundwater areas as well as water quantity and quality are known based on previous hydrological studies. The average Kitigan Zibi household has approximately five occupants (Kitigan Zibi FN, personal communication, 5 April, 2012). The province of Québec recommends a well-water quantity of 750–885 litres per hour for a household of this size (MDDEP, 2002c). Although individual wells produce approximately 568–758 litres per house per hour, water quantity is adequate for household use (Kitigan Zibi FN, personal communication, 30 April, 2010). However, water quality is a concern (Kitigan Zibi FN, personal communication, 30 April, 2010). In 1999, a Health Canada study found high levels of uranium, a toxic heavy metal, present in groundwater and issued a 'do not consume' drinking water advisory for well water users (Harden and Levalliant, 2008).

**Figure 2** Location of the Kitigan Zibi First Nation in the Gatineau River Watershed

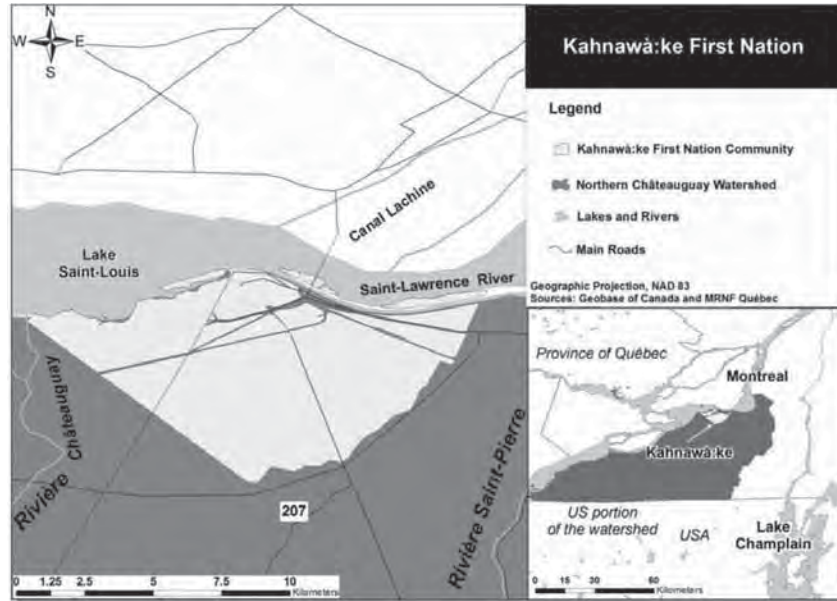
## 2.2 Kahnawà:ke First Nation

Kahnawà:ke is an urban community of 7,556 Mohawk residents, located 10 km southwest of Montréal on the south shore of Lake Saint-Louis (Fig. 1). The community covers 4,811 hectares (11,888 acres) and is part of the Châteauguay River Watershed (Figures 1 and 3; INAC, 2010a). There is year-round road access and it is located less than 50 km from the nearest service centre (INAC, 2010a).

Kahnawà:ke relies on surface water from the St. Lawrence for the larger part of its drinking water needs, and the community has over 60 years of experience in drinking water management. Piped water distribution was installed in the 1950s to core homes and now the piped water distribution network extends to 2,200 households and 114 businesses (Kahnawà:ke FN, personal communication, 5 October, 2010). Some 300 households and 49 businesses rely on well water for non-drinking purposes and individual wells are not monitored frequently (Kahnawà:ke FN, personal communication, 25 May, 2010). Three to four individual wells per week are monitored, which means an individual well may be tested only once every two years (Kahnawà:ke FN, personal communication, 30 April, 2010).

## 3 Capacity in Integrated Water Resources Management

The United Nations Development Program (UNDP) (1997) defines capacity as the “ability of individuals, institutions and societies to perform functions, solve problems and set and achieve objectives.” Since the early 1990s, capacity in the context of water management has

**Figure 3** Location of the Kahnawà:ke First Nation in the Châteauguay Watershed

been viewed as critical (Hartvelt and Okun, 1991; Franks, 1999). The nature of an integrated approach to water resources management also requires capacity across various areas and recognises interdependencies (Franks, 1999; Cap-Net, 2006). The UNDP recognises that capacity building is a long-term, continuing process and is a strategic element for the sustainable management of the water sector (Biswas, 1996; Franks, 1999).

The UN-Water Decade Program on Capacity Development (UNW-DPC) prioritises capacity development activities and requires capacity needs assessment and gap analysis, as well as the implementation of innovative capacity development methodologies (UN Water, 2009). *A Handbook for IWRM in Basins* suggests a successful basin management strategy anticipates the need to strengthen capacity and fund capacity building, and basin organisations are encouraged to develop programmes to build capacity (GWP, 2009). Furthermore, it is suggested that capacity development is necessary for disenfranchised groups to ensure their involvement in planning and implementation (GWP, 2009). Essentially, capacity is an enabler and driver of IWRM (Van der Zaag, 2005) and the water sector is highly dependent on individual and institutional capacities (Blokland et al., 2009). Strengthening capacity is an integral component of IWRM.

#### 4 First Nation capacity challenges in Canada

FNs in Canada experience a wide spectrum of capacity challenges related to water resources management. With respect to human resources capacity, FN water treatment operators are critical to the delivery of safe drinking water (O'Connor, 2002; Swain et al., 2006); however, FN communities often lack certified or qualified personnel to operate water treatment facilities, which results in considerable risk and problems (INAC, 2003; Smith et al., 2006). Despite

the fact that water treatment operators are critical for the delivery of safe drinking water, operator training certification and retention of qualified individuals are also major issues in FN communities (OAG, 2005; Smith et al., 2006).

Financial capacity is another area of concern. FN communities rely heavily on Indian and Northern Affairs Canada (INAC) for capital and operational funding and Health Canada's FNs and Inuit Health Branch for monitoring drinking water quality (Smith et al., 2006). Despite substantial funding aimed at addressing water quality in FN communities, the efforts of the federal government have yielded limited improvement in drinking water (OAG, 2005). In addition to a community's level of poverty, the overall direct impact of a community's ability to finance the operation and maintenance (O & M) of water treatment facilities is limited. FNs are responsible for 20% of O & M costs for water systems, a heavy financial burden in communities with high unemployment and little likelihood of recovering costs from the community (Chiefs of Ontario, 2001; OAG, 2005; Swain et al., 2006). In addition, communities can seldom afford operator-training expenses when presented with financial constraints (Swain et al., 2006). Despite FN financial challenges, communities are required to meet the same health-based water quality standards as larger drinking water systems, even if they lack resources and economies of scale that larger systems enjoy (Smith et al., 2006). Limited financial resources place a major risk on O & M objectives, thereby jeopardising safe drinking water in FN communities.

Technical capacity in more remote or smaller FN communities is a problem, as these communities have greater difficulty than non-remote or larger communities in coping with technical and managerial challenges specific to water-related activities (Morris et al., 2007; Hrudey, 2008)<sup>3</sup>. Some 43% of aboriginal people live in remote communities or settlements and comprise 30% of Northern Canada's population (MacLeod et al., 1998). Smaller systems can contribute to marginalised water infrastructures, and in situations where there is a complete absence of water infrastructure, this can lead to the transmission of diseases (Health Canada, 2005). Among 61 FN communities in Manitoba, those without wells or running water accounted for 89% of the *Shigella* cases in the early 1990s (Clarke et al., 2002). The reported incidence of *Shigella* among FNs communities (74.1 per 100,000 individuals) was 26 times greater than that of their non-aboriginal Canadian counterparts (2.8 per 100,000 individuals) (Clarke et al., 2002).

The 2001 National Assessment of Water and Wastewater Systems in FN communities conducted by INAC was based on an on-site inspection of all FN water systems and included an evaluation of system performance, associated risk levels and operating practises (INAC, 2003). Of 740 community water systems, 46% were classified as posing a medium water quality risk and 29% were classified as posing potentially high risks (INAC, 2003). High and medium water quality risk assessments occurred as a result of a failure to meet one or more Maximum Acceptable Concentration (MAC) parameters<sup>4</sup>. Failure to meet MAC parameters can occur due to a lack of: regular testing procedures, records maintenance, operator knowledge in how to run the water system, emergency procedures, safety equipment and operating manuals in the facilities. Poor raw water sources, inadequate treatment, equipment failure and absence of backup equipment or power sources can also result in a failure to meet MAC standards (INAC, 2003). In 2003, the federal government responded to these poor results with the FNs Water Management Strategy (FNWMS), which was afforded a \$600 million budget to improve water and wastewater systems in FN communities (INAC, 2004; OAG, 2005). In 2005, the Office of the Auditor General audited the FNWMS and concluded that residents of FN communities did not benefit from the same level of drinking water protection as other Canadian communities (OAG, 2005).



In 2006, the Plan of Action for Drinking Water in FN communities yielded the *Protocol for Safe Drinking Water for FN Communities*. On-going training of 875 operators through the Circuit Rider Training Program and 24-hour access to a support hotline, led to a decrease in high-risk drinking water systems from 193 to 97 (INAC, 2007). Building on progress under the Plan of Action for Drinking Water in FN Communities, the 2008 FNs Water and Wastewater Action Plan (FNWWAP) received \$330 million in funding. There were 49 high-risk drinking water systems identified, significantly below the 193 identified in 2006 (INAC, 2010d). In 2009, the Economic Action Plan (EAP) targeted \$165 million for water and wastewater infrastructure projects in 18 FNs communities across the country (INAC, 2010c).

Institutional capacity refers to a water regulatory regime that provides rules and standards to ensure water quality and safety. This type of capacity is generally absent in Canadian FN communities. Currently, drinking water safety in FN communities is managed through a series of guidelines, protocols and contracts between Indian and Northern Affairs Canada (INAC) and FN communities (OAG, 2005; Swain et al., 2006; Duncan and Bowden, 2009; MacIntosh, 2009). There is a general consensus among senate committees, independent commissions and political representatives like the Assembly of FNs (AFN), that the current institutional situation produces unacceptable levels of risk to public health and that a regulatory framework is needed for FNs (OAG, 2005; Swain et al., 2006; Duncan and Bowden, 2009; MacIntosh, 2009). The Office of the Auditor General (2005) report concluded that the federal government's passing fiduciary and water provisioning responsibilities to FNs creates confusion in regards to where the ultimate responsibility falls. In 2005, when over 800 members of Kashechewan FNs were evacuated after *E. coli* was discovered in their water supply, responsibility shifted amongst the federal, provincial governments and Kashechewan FNs. When water pathogens are detected, there is no clear protocol on how to proceed or who to assign responsibility to.

Another aspect of institutional capacity is jurisdiction. The limits and powers of a territory are of great concern in integrated water management processes. A watershed protection report submitted to the Ministry of Environment of Ontario cited three non-aboriginal municipalities that clearly lacked municipal authority "to address threats to vulnerable drinking water sources in existing built-up areas and from existing activities" (Hill et al., 2009). Commissioner O'Connor's Report on the Walkerton Inquiry recommended working toward intergovernmental coordination, particularly with representatives of Fisheries and Oceans Canada, Environment Canada, Indian and Northern Affairs Canada and Agriculture and Agri-Food Canada. However, O'Connor (2002) cautioned that this is complex "in an area where constitutional jurisdiction is not always clear." If non-aboriginal municipalities are subject to a lack of authority as it relates to watershed management, then FN communities will likely be subject to even greater governance complexities. Another aspect of institutional capacity relates to customary water rights. Aboriginal peoples have water rights, unless limited or properly extinguished (Phare, 2006). Section 35 of the *Canadian Constitution Act of 1982* affirms and protects aboriginal rights to occupy land, fish, hunt, trap and generally use goods produced by the land and water (Craig, 2003; Kempton, 2005; Phare, 2006). Prior to 1982, only the federal government (and not provincial governments) could extinguish aboriginal and treaty rights, whereas today, neither government can extinguish water rights (Kempton, 2005).

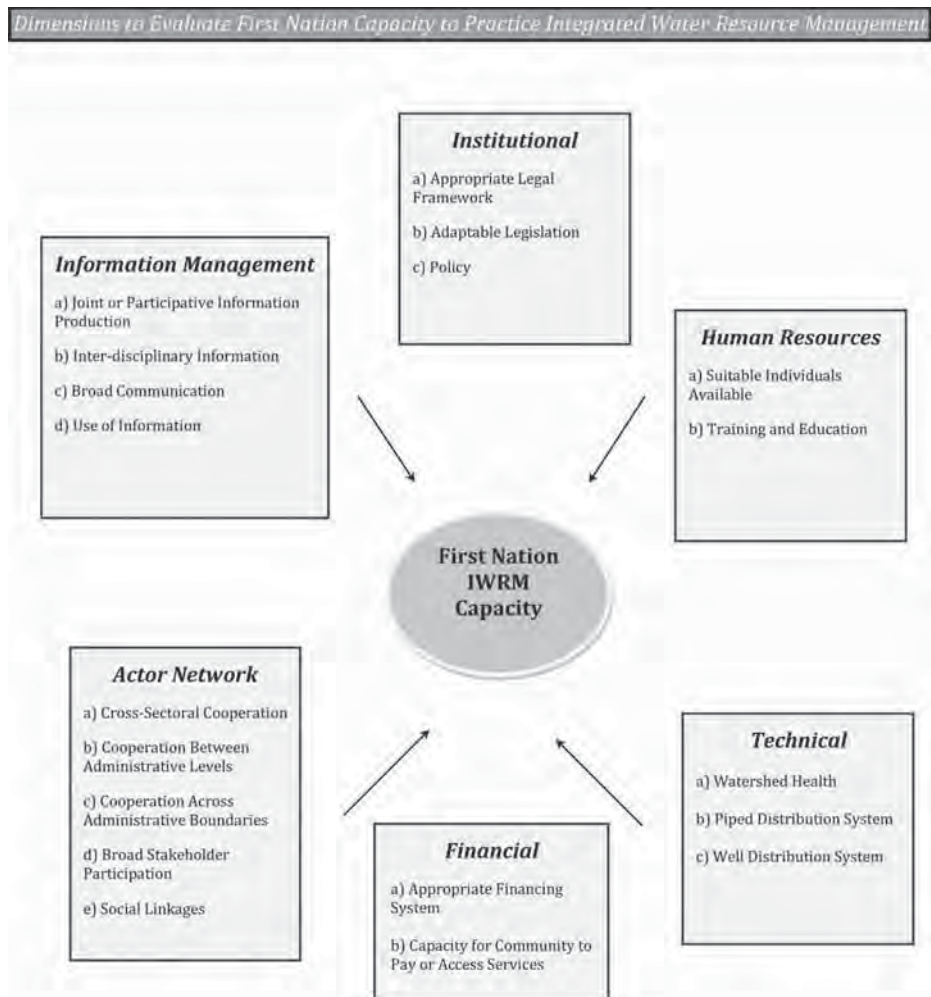
## **5 Applying an analytical framework to evaluate community capacity to practise IWRM**

For the purpose of this study, six dimensions and a number of sub-dimensions were employed in evaluating capacity (Figure 4). Each sub-dimension is composed of various capacity



indicators that characterise the dimension. The selection of capacity dimensions was based on previous research in the literature that demonstrated the necessity of a given capacity as it related to an aspect of IWRM. The capacity dimensions employed include: actor network, information management, human resources, technical, financial and institutional capacity. Capacity indicators provided a metric for identifying trends toward or away from an intended objective. To achieve the study’s objectives, 79 indicators were drawn from the literature specifically related to watershed management, source water protection, drinking water management and community capacity (Pahl-Wostl et al., 2007; Timmer et al., 2007; Cap-Net, 2008; De Carvalho et al., 2008; Raadgever et al., 2008). As IWRM continues to evolve, future efforts need to include indicators identified and developed by Indigenous communities, particularly of community-based knowledge specific to environmental resources more commonly referred to as ‘Traditional Ecological Knowledge’ (TEK).

**Figure 4** Conceptual framework to evaluate First Nation capacity to practise IWRM



A four-level rating scheme was adapted from a 'Summary Indicator Table' developed by the Environmental Finance Center's (2005) assessment of drinking water safety as it relates to financial capacity. Based on qualitative data derived from interviews, response content was evaluated as either positive or negative. Each capacity indicator was evaluated as having an 'absence' of capacity if it did not meet or partially met the requirements, or a 'presence' of capacity if it met or would meet the requirements in the future. In the assessment of each capacity dimension, all indicators were weighted equally, as suggested by McGuire et al. (1994). Subsequently, an overall designation for each capacity dimension was denoted as either having a capacity 'presence', or 'absence', while a 'partial' capacity indicated a dimension that partly meets the criteria.

Primary data for this study were derived from one-on-one interviews in the two participant FN communities. In total, five interviews were conducted in Kitigan Zibi and three in Kahnawà:ke FN. Interviews were conducted with employees and leaders with expertise and experience in natural resources management, water system operations, water testing, engineering, finance and governance. Individuals included Chiefs, forest and civil engineers, water treatment operators, public work directors and environmental health and safety technicians, as well as environmental and financial administrators. The names, training and jobs of interviewees were not disclosed in any part of this research. Total anonymity was necessary to build an enabling environment where all interviewees felt there were no consequences for their opinions, observations and experiences. This is particularly important in smaller communities where identifying participants by profession easily identifies the interviewee. Qualitative data were gathered by using both structured interviews, consisting of predetermined questions, with the same question order and wording (Kumar, 2005), as well as questions modified to incorporate flexibility and allow exploration of emerging information (Kumar, 2005). The presence and absence of capacity indicators for each dimension were recorded, based on interviews, in the *FN Capacity to Practise IWRM Indicator Ratings Table* (Tables 1 to 7). At a later time, the results were shared with each interviewee for verification that their input was correctly evaluated to reduce research error.

**Table 1** Summary of capacity results for Kitigan Zibi and Kahnawà:ke First Nations

Capacity type	Total indices	Capacity indicators (Present:total)	
		Kitigan Zibi	Kahnawà:ke
Actor network	18	2:18	6:18
Information management	10	4:6	4:6
Human resources	4	2:4	3:4
Technical	26	14:26	12:26
Financial	12	4:12	3:12
Institutional	9	6:9	8:9

## 6 Results and discussion

The results represent findings for case studies of two FN communities in the province of Québec and should not be generalised to all FN communities in Canada. Rather, the findings of this research are an initial discussion on FN capacity along with an identification of capacity preconditions required for their participation in IWRM.

**Table 2** Actor network capacity indicator ratings, showing whether a particular capacity is not met, partially met, met, or met and expected to be met in the future

<i>Elements of capacity</i>	<i>Capacity met</i>							
	<i>Kitigan Zibi</i>				<i>Kahnawà:ke</i>			
	<i>No</i>	<i>In part</i>	<i>Yes</i>	<i>Yes + future</i>	<i>No</i>	<i>In part</i>	<i>Yes</i>	<i>Yes + future</i>
<i>Cross-Sectoral Cooperation</i>								
• Partnerships with different communities & stakeholders	•						•	
• Conflicts with other parties (communities, stakeholders) dealt with constructively, resulting in inclusive agreements to which the parties are committed	•					•		
• Use of cross-sectoral analysis to identify emergent problems and for policy implementation	•					•		
<i>Cooperation Between Administrative Levels</i>								
• FN governments involved in decision-making processes with the federal departments (vertical linkages)			•				•	
• Conflicts dealt with constructively, resulting in inclusive agreements to which the parties are committed			•				•	
<i>Cooperation Across Administrative Boundaries</i>								
• Downstream communities involved in decision making by upstream communities	•					•		
• FN community part of a cooperation structure (e.g., watershed associations)	•					•		
• Conflicts dealt with constructively, resulting in inclusive agreements to which the parties are committed	•						•	
<i>Broad Stakeholder Participation</i>								
• Legal provisions concerning access to information, participation in decision-making (e.g., consultation requirements)	•					•		

**Table 2** Actor network capacity indicator ratings, showing whether a particular capacity is not met, partially met, met, or met and expected to be met in the future (continued)

<i>Elements of capacity</i>	<i>Capacity met</i>							
	<i>Kitigan Zibi</i>				<i>Kahnawà:ke</i>			
	<i>No</i>	<i>In part</i>	<i>Yes</i>	<i>Yes + future</i>	<i>No</i>	<i>In part</i>	<i>Yes</i>	<i>Yes + future</i>
• Community includes cooperation structures from non-government groups	•				•			
• Community contributes to agenda setting, analysing problems, developing solutions and making decisions at the watershed scale		•				•		
• Community undertakes parts of watershed management themselves, e.g., through watershed associations	•				•			
• Federal Government takes community input seriously		•					•	
• Provincial Government takes community input seriously		•					•	
• Tribal Council or Environmental department takes community input seriously				•			•	
<i>Social Linkages</i>								
• Clear leadership for water protection at the watershed level exists	•					•		
• Community members have awareness and support for watershed protection				•			•	
• Community members regularly involved in decisions as it pertains to drinking water management and environmental protection		•					•	

### 6.1 Actor network capacity

This type of capacity requires cooperation, communication and exchange of information. Hence, the need for partnerships is essential in collaborative water management efforts (Yillia et al., 2003). Capacity is improved when stakeholders coordinate, facilitate and maintain active linkages to provide vision and direction (De Loë et al., 2002) and ultimately, partnerships can help overcome the 'silo' effect (Mitchell, 2006).

**Table 3** Information management capacity indicator ratings, showing whether a particular capacity is not met, partially met, met, or met and expected to be met in the future

<i>Elements of capacity</i>	<i>Capacity met</i>							
	<i>Kitigan Zibi</i>				<i>Kahnawà:ke</i>			
	<i>No</i>	<i>In part</i>	<i>Yes</i>	<i>Yes + future</i>	<i>No</i>	<i>In part</i>	<i>Yes</i>	<i>Yes + future</i>
<i>Joint or Participative Information Production</i>								
<ul style="list-style-type: none"> <li>Community is involved in setting the terms of reference and supervising the research, or is at least consulted (interviews, surveys, etc.) at the watershed scale</li> </ul>		•					•	
<i>Interdisciplinary</i>								
<ul style="list-style-type: none"> <li>Different disciplines involved in defining and executing research, in addition to technical and engineering sciences, also include ecology, social sciences, etc.</li> </ul>				•			•	
<i>Broad Communication</i>								
<ul style="list-style-type: none"> <li>Different levels of governments exchange information and data with other governments (federal, tribal councils, band councils)</li> </ul>				•			•	
<ul style="list-style-type: none"> <li>Community actively disseminates information and data to the public (internet, literature, brochures, media, etc.)</li> </ul>		•				•		
<i>Use of Information</i>								
<ul style="list-style-type: none"> <li>New information used in public debates</li> </ul>				•		•		
<ul style="list-style-type: none"> <li>New information influences federal policy</li> </ul>		•				•		
<ul style="list-style-type: none"> <li>Data needed to manage water supplies, delineate watersheds and aquifers and develop source protection plans available</li> </ul>				•			•	
<ul style="list-style-type: none"> <li>Community monitors and collects data (e.g., produces quality data)</li> </ul>				•			•	



**Table 3** Information management capacity indicator ratings, showing whether a particular capacity is not met, partially met, met, or met and expected to be met in the future (continued)

Elements of capacity	Capacity met							
	Kitigan Zibi				Kahnawà:ke			
	No	In part	Yes	Yes + future	No	In part	Yes	Yes + future
• Water management information available to managers and other stakeholders if requested			•				•	
• Comprehensive understanding achieved with open, shared information sources that fill gaps and facilitate integration at the watershed level		•				•		

In this study, 18 actor network capacity indicators were employed (Tables 1, 2), including vertical and horizontal linkages that encompass cross-sectoral cooperation, cooperation between administrative levels, across administrative boundaries and broad stakeholder participation, as adopted in the study by Raadgever et al. (2008) on water management regimes. Another important indicator included in this study is cross-sectoral analysis to identify emerging problems and integrate policy implementation (Pahl-Wostl et al., 2007). Social linkages, a vital element of actor network, include indicators identified as specific to clear leadership for water protection at the watershed level, in addition to community awareness and support for watershed protection (Timmer et al., 2007). Another key aspect of social linkages was developed by Lebel (2008) and evaluates community members’ involvement in decisions pertaining to drinking water management and environmental protection. The actor network capacity indicators selected for this study were chosen to reflect the diverse vertical and horizontal linkages that exist in drinking water management, environmental protection and watershed participation.

A clear absence of actor network capacity was apparent for the Kitigan Zibi FN (Tables 1, 2). The reasons were two-fold:

- a lack of partnerships with neighbouring communities
- a lack of participation in watershed associations.

In the first instance, interviewees indicated a non-aboriginal community was situated illegally on Algonquin territory. FNs territorial claims are problematic in Quebec, as Quebec’s identity comes with a “profound sense of belonging to the territory traditionally recognized by cartographers” (Salée, 1995). However, FN identity to ancestral land is profound and is in direct conflict with Quebec’s identity (Kitigan Zibi FN, personal communication, 30 April, 2010). Furthermore, an interviewee suggested the neighbouring municipality is “*threatened by the economic rise of Kitigan Zibi*” and that socio-political tensions exist as a result. In the second instance, a lack of participation in watershed associations can be attributed to this Nation’s strong preference to be recognised as an independent nation with government status and not just as another ordinary stakeholder (Kitigan Zibi FN, personal communication,

**Table 4** Human resources capacity indicator ratings, showing whether a particular capacity is not met, partially met, met, or met and expected to be met in the future

<i>Elements of capacity</i>	<i>Capacity met</i>							
	<i>Kitigan Zibi</i>				<i>Kahnawà:ke</i>			
	<i>No</i>	<i>In part</i>	<i>Yes</i>	<i>Yes + future</i>	<i>No</i>	<i>In part</i>	<i>Yes</i>	<i>Yes + future</i>
<ul style="list-style-type: none"> <li>• <i>Availability of Suitable Employees</i></li> <li>• Sufficient number of employees dedicated to water management, environmental protection or rights-holder participation</li> <li>• Access to individuals with the appropriate level of education and expertise to adequately support water management, environmental protection or rights-holder participation</li> </ul>		•		•		•	•	
<i>Training and Education</i>								
<ul style="list-style-type: none"> <li>• Education and training opportunities available to staff members to participate in, helping them to contribute to water management, environmental protection or rights-holder participation activities</li> <li>• Education and training opportunities regularly taken up by staff members from various departments to participate and contribute to water management, environmental protection or rights-holder participation activities</li> </ul>			•				•	

30 April, 2010). In Canada, aboriginal people’s customary right to water ensures a rights-holder status, thereby placing them in a unique position unlike that of any other stakeholder (Phare, 2006). There is dissatisfaction with forestry and mining industry consultations, which are viewed as documents outlining decisions that have already been made either without their input or which ignore the concerns they have voiced (Kitigan Zibi FN, personal communication, 30 April, 2010). As such, the potential opportunity for collaboration with other actors in a watershed association may produce the same dissatisfaction experienced with government or industry (Kitigan Zibi FN, personal communication, 30 April, 2010).

A similar absence of actor network capacity in the Kahnawà:ke FN is attributable to multiple factors. An absence is due to previous political conflict in Oka, a preference

**Table 5** Technical capacity indicator ratings, showing whether a particular capacity is not met, partially met, met, or met and expected to be met in the future

Elements of capacity	Capacity met							
	Kitigan Zibi				Kahnawà:ke			
	No	In part	Yes	Yes + future	No	In part	Yes	Yes + future
<i>Watershed Health</i>								
• Community drinking water quality monitored (throughout the watershed) regularly (daily tests)	•				•			
• Community drinking water quality monitored (throughout the watershed) regularly (weekly and quarterly tests)	•				•			
• Community drinking water quality monitored (throughout the watershed) regularly (annual tests)	•						•	
• Community groundwater recharge areas are identified			•				•	
• Community source water areas incorporated into official plans			•				•	
• Potential water supply contaminant sources (point & non-point) identified		•				•		
<i>Piped Distribution System</i>								
• Community drinking water quality meets established drinking water standards			•				•	
• Community drinking water quality monitored (within the water distribution system) regularly (daily tests)			•				•	
• Community drinking water quality monitored (within the water distribution system) regularly (weekly and quarterly tests)			•				•	
• Community drinking water quality monitored (within the water distribution system) regularly (annual tests)			•				•	
• Community groundwater recharge areas are identified			•		•			
• Community source water areas incorporated in official plans			•		•			

**Table 5** Technical capacity indicator ratings, showing whether a particular capacity is not met, partially met, met, or met and expected to be met in the future (continued)

<i>Elements of capacity</i>	<i>Capacity met</i>							
	<i>Kitigan Zibi</i>				<i>Kahnawà:ke</i>			
	<i>No</i>	<i>In part</i>	<i>Yes</i>	<i>Yes + future</i>	<i>No</i>	<i>In part</i>	<i>Yes</i>	<i>Yes + future</i>
• Potential water supply contaminant sources (point & non-point) identified			•			•		
• Physical infrastructure adequate to produce safe drinking water for community residents			•				•	
• Physical infrastructure adequate to distribute safe drinking water for community residents			•				•	
• Source water adequate in terms of quantity			•				•	
• Source water adequate in terms of quality	•					•		
<i>Well Distribution System</i>								
• Community drinking water quality meets established drinking water standards		•				•		
• Community drinking water quality monitored (within the water supply and distribution system) regularly (daily tests)	•				•			
• Community drinking water quality monitored (within the water supply and distribution system) regularly (weekly and quarterly tests)		•				•		
• Community drinking water quality monitored (within the water supply and distribution system) regularly (annual tests)			•				•	
• Potential water supply contaminant sources (point & non-point) identified		•				•		
• Physical infrastructure adequate to produce safe drinking water for the community's residents		•				•		
• Physical infrastructure adequate to distribute safe drinking water for the community's residents		•				•		

**Table 5** Technical capacity indicator ratings, showing whether a particular capacity is not met, partially met, met, or met and expected to be met in the future (continued)

Elements of capacity	Capacity met							
	Kitigan Zibi				Kahnawà:ke			
	No	In part	Yes	Yes + future	No	In part	Yes	Yes + future
• Source water adequate in terms of quantity			•				•	
• Source water adequate in terms of quality	•					•		

for aboriginal sovereignty and a lack of participation in local watershed associations (Tables 1, 2). However, it was suggested that previous political conflicts have led the federal government to be more sensitive to ensuring conflicts are resolved (Kahnawà:ke FN, personal communication, 25 May, 2010). In addition, Kahnawà:ke has rejected opportunities to integrate with the province and prefers to promote aboriginal sovereignty (Alfred, 1995). French, the operative language for the local watershed association, is considered a barrier for the participation of Kahnawà:ke, as, in general, residents speak only English and Mohawk. An interviewee thought “*their participation [in the francophone watershed association] required more time for translations*” and this limited their role in local watershed meetings, thereby hindering their participation (Kahnawà:ke FN, personal communication, 26 May, 2010). Overall, the socio-political realities of each community greatly determine its ability to work collaboratively to identify common water resource concerns.

### 6.2 Information management capacity

This capacity, as it relates to integrated water management, requires that information be shared and generated collaboratively. Cooperation in information management helps develop trust and collaboration amongst stakeholders. Information has to be accessible, shared and integrated to enable decision-making (Kennedy et al., 2009). Information management fosters greater technical capacity, mutual understanding and shared insights (Mostert et al., 1999; Van der Zaag and Savenije, 2000).

In Kitigan Zibi, FN there exists a partial information management capacity (Tables 1, 3). Although the community has expertise (or the ability to locate a third party) in producing information at the local level, there is a clear lack of joint or participative information produced with partners at the watershed scale. In IWRM, it is the sharing of information that is imperative to developing information management capacity (Mostert et al., 1999; Van der Zaag and Savenije, 2000; Raadgever et al., 2008; Kennedy et al., 2009). This lack of information management is primarily due to Kitigan Zibi’s lack of participation in the watershed association, poor regional partnerships with other municipalities and dissatisfaction with consultation processes with government or private industries.

In the Kahnawà:ke FN there is a presence of information management capacity (Tables 1, 3). Kahnawà:ke can produce information on a joint or participative-basis, use



**Table 6** Financial capacity indicator ratings, showing whether a particular capacity is not met, partially met, met, or met and expected to be met in the future

<i>Elements of capacity</i>	<i>Capacity met</i>							
	<i>Kitigan Zibi</i>				<i>Kahnawà:ke</i>			
	<i>No</i>	<i>In part</i>	<i>Yes</i>	<i>Yes + future</i>	<i>No</i>	<i>In part</i>	<i>Yes</i>	<i>Yes + future</i>
<i>Appropriate Financing System</i>								
• Sufficient (public and private) resources available for water management initiatives (e.g. source water protection, watershed participation, infrastructure, water system projects)	•							•
• Costs recovered from the users by public and private financial instruments to maintain a balanced budget			•					•
• Decision-making and financial control under the same leadership			•					•
• Water rates reflect the cost of protecting and providing drinking water (including treatment, distribution, maintenance and source water protection)	•				•			
• Able to obtain funding from outside the community		•						•
• Able to obtain funding from inside the community		•						•
• Stable funding	•				•			
• Funding surpluses saved for future water projects			•					•
<i>Capacity for a Community to Pay or Access Services</i>								
• What level of education do most community members have		•						•
• Unemployment rate		•						•
• Average income level		•						•
• Work days lost per annum due to water related diseases			•					•

different disciplines in research, share information and they have water management data and the ability to monitor and collect data (Kahnawà:ke FN, personal communication, 26 May, 2010). Since Kahnawà:ke does not participate in watershed associations, a comprehensive understanding could not be achieved because information was not shared to fill gaps and facilitate integration as required for watershed management in a collaborative setting.

**Table 7** Institutional capacity indicator ratings

<i>Elements of capacity</i>	<i>Capacity met</i>							
	<i>Kitigan Zibi</i>				<i>Kahnawà:ke</i>			
	<i>No</i>	<i>In part</i>	<i>Yes</i>	<i>Yes + future</i>	<i>No</i>	<i>In part</i>	<i>Yes</i>	<i>Yes + future</i>
<i>Legal Aspects</i>								
<i>— Appropriate Legal Framework —</i>								
• There are complete and clear legal frameworks for water management (with sufficient detail)		•					•	
<i>— Adaptable Legislation —</i>								
• Federal laws and regulations easily changed	•					•		
<i>— Actual Implementation of Policies —</i>								
• Plans and policies actually implemented			•				•	
• Local policies reviewed and changed periodically		•					•	
• Policies are flexible and not rigid when there are good reasons not to implement them (e.g., new and unforeseen circumstances and new insights)			•				•	
<i>— Planning —</i>								
• There are community planning strategies and by-laws that protect current drinking water supplies			•				•	
• Land use activities controlled in community				•			•	
• well fields, recharge and watershed water supply areas								•
<i>— Long Term Horizon —</i>								
• Solutions for short-term problems which do not cause more problems in the (far) future (20 years or more)			•				•	
• Preparations being made for the (far) future (20 years or more)			•				•	

Both participant communities indicate a high level of capacity as it relates to producing quality water data. However, a lack of participation in watershed associations hinders the distribution of information to other stakeholders. It is apparent from this study that information management capacity has a direct relationship with actor network capacity. If

there is a presence of an actor network capacity, there is a greater likelihood of information management capacity as it relates to sharing information. Trust and cooperation are necessary components to ensure information is accessible, shared and integrated to enable decision-making at the watershed level.

### *6.3 Human resources capacity*

This capacity refers to the education and training individuals currently possess in water management, protection or rights-holder participation activities and also includes opportunities for continued professional growth. Human resources capacity is necessary for competent water management (Forster, 1997) and essentially, links education, training and the abilities of individuals to achieve sustainable water stewardship. Regional capacity and human resources development are important elements in IWRM (Forster, 1997; Van Der Zaag, 2003; Gumbo et al., 2005).

In Kitigan Zibi FN there exists a partial human resources capacity (Tables 1, 4). Although Kitigan Zibi has a sufficient number of employees for O & M, interviewees indicated a dedicated staff person to exclusively manage watershed health, environmental protection where rights-holder participation was required (Kitigan Zibi FN, personal communication, 30 April, 2010). Limited financial resources are a concern when trying to ensure there is a dedicated staff person to manage watershed issues. Although there are education and training opportunities available to staff members to participate in and later contribute to water management, environmental protection or rights-holder participation, the onus is on the individual to take these opportunities (Kitigan Zibi FN, personal communication, 30 April, 2010). Furthermore, these opportunities are only available if additional funding is located, which is not an easy endeavour, particularly in the case of Kitigan Zibi, which operates on a zero debt policy (Kitigan Zibi FN, personal communication, 25 May, 2010).

In the Kahnawà:ke FN, human resources capacity is present (Tables 1, 4). In regards to a sufficient number of employees, the interviewees again voiced the same concerns and preferred a dedicated staff person to exclusively manage watershed health, environmental protection and rights-holder participation. An interviewee in Kahnawà:ke thought financial resources have become more constrained in recent years and that this hinders the possibility of hiring a staff person exclusively for watershed-related activities (Kahnawà:ke FN, personal communication, 25 May, 2010). Financial capacity has a direct impact on human resources capacity. Without adequate funds, staff to support drinking water management and environmental protection is not possible. If financial resources are limited, then priority is directed to drinking water management, which thereby takes precedence over activities associated with watershed health.

Based on this study, limited financial resources have been found to greatly hinder the ability to hire staff to exclusively manage watershed matters. Current staff cannot be expected to adequately manage watershed responsibilities, in addition to their present responsibilities. A staff person to support integrated water management matters would directly benefit information management, actor network and technical capacity.

### *6.4 Technical capacity*

This capacity encompasses watershed health, as well as piped and well water distribution systems. Kitigan Zibi FN showed partial technical capacity (Tables 1 and 5). With

respect to watershed health, there exists an absence of capacity, primarily due to lack of monitoring of water quality throughout the watershed and to not knowing where potential water supply contaminant sources are. Limited human and financial resources impact overall watershed health activities (Kitigan Zibi FN, personal communication, 30 April, 2010).

With respect to the piped distribution system, there exists a technical capacity, reflected by water quality meeting established drinking water standards, regular water monitoring, knowledge of groundwater recharge areas and potential water supply contaminants (point and non-point source), adequate physical infrastructure to distribute and produce safe drinking water, incorporation of water sources into plans and adequate source water quantity. In 2010, Kitigan Zibi had their first water and wastewater system approved, at a cost of \$10 million obtained from the EAP. It was completed in March 2011, and presently 195 households (or 37% of households) are connected to a piped water distribution system (INAC, 2009b, INAC, 2011). Currently, Kitigan Zibi is self-sufficient in providing water and wastewater services to 236 households (or 45% of households) (Kitigan Zibi FN, personal communication, 30 April, 2010). The Band Council hopes to deliver piped water to the remaining 289 households within the next 5–10 years (Kitigan Zibi FN, personal communication, 30 April, 2010). Interviewees were highly in favour of piped distribution as a means of self-sufficiency and this installation is another step towards achieving this goal. It should be noted that EAP funds were distributed to only 18 FN communities across Canada, including three in Québec, which include both Kitigan Zibi and Kahnawà:ke.

With respect to well distribution systems, the results clearly indicate an absence of capacity. Groundwater sources contain uranium, which contributes to poor water quality. In addition, the monitoring of wells is time-intensive and the number of wells to be monitored greater than available human resources can handle.

The Kahnawà:ke FN shows partial technical capacity (Tables 1, 5). In terms of watershed health there is an absence of capacity primarily due to lack of water quality monitoring throughout the watershed and lack of knowledge about where potential water supply contaminant sources are (point and non-point source). In a manner similar to Kitigan Zibi, limited human and financial resources to support watershed health activities are a reason for this capacity deficiency.

In terms of piped distribution systems, there is a presence of capacity. The following indicators support piped distribution capacity: the presence of water quality that meets established drinking water standards, regular water monitoring, adequate physical infrastructure to distribute and produce safe drinking water and adequate source water quantity.

In terms of well distribution systems, there is an absence of capacity. The following indicators do not support well distribution system capacity: drinking water does not meet established standards, monitoring of wells is not done regularly, there is poor physical infrastructure to distribute or produce safe drinking water and the source water is inadequate in quality.

In summary, both communities showed a partial capacity with regard to watershed health, a presence of capacity in piped distribution systems and an absence of capacity for well distribution systems. It should be noted that a lack of information or knowledge results in an absence of capacity indicators, which in turn contributes to information gaps related to problem solving within an IWRM context and with regards to achieving safe drinking water or environmental sustainability.

### 6.5 Financial capacity

This capacity represents the ability to access, generate and save funds for drinking water and environmental stewardship.

Results for the Kitigan Zibi FN indicate an absence of financial capacity (Tables 1, 6). Vulnerabilities in internal and external funding sources contribute to overall poor financial capacity. FNs rely heavily on INAC for external funds. In January 2009, the federal government administered funds of \$165 million for water and wastewater projects for aboriginal peoples under the EAP (INAC, 2009a; INAC, 2009b). The 2010 EAP approval of Kitigan Zibi's first water and wastewater system initially did not include a wastewater system. The community encouraged the installation of both systems and did not sign the agreement until provisions were made to include a wastewater system (Kitigan Zibi FN, personal communication, 30 April, 2010). An interviewee noted that there was a constant struggle to secure external funds from INAC and that funds were not always accessible (Kitigan Zibi FN, personal communication, 30 April, 2010).

In addition, the community's inability to pay or access services produces internal financial vulnerabilities. In general, community members find seasonal employment in the forestry and mining industries, but experience challenges in securing long-term employment (for those not employed by the Band office) (Kitigan Zibi FN, Personal Communication, 30 April, 2010). The economic implications of these community employment trends do not support a healthy base for the Band office to rely on for financing water-related activities.

The Kitigan Zibi community indicated that residents do not pay what is required to fund drinking water services, as current fees for water services are substantially lower than the real cost of operating and maintaining water facilities and services, particularly when taking into account rising input costs. This low capacity for communities to financially contribute exerts great financial strain on limited Band resources. In Kitigan Zibi, the Band Office offers insurance to households for maintenance and repairs for well distribution and septic systems. The sum of \$85/household insures an individual well and an additional \$85/household insures septic systems (Kitigan Zibi FN, Personal Communication, 17 September, 2010). This is voluntary and only 60% of the community pays it.

An absence of financial capacity heavily impacts activities associated with watershed health. There is a heavy dependence on external funding and there are no financial sources generated from the community or secured with external government funding specified for watershed health. An interviewee specified financial resources are required to attend meetings and to collect and maintain data necessary for participation and without financial resources, watershed participation is not a viable option (Kitigan Zibi FN, personal communication, 30 April, 2010).

In the case of the Kahnawà:ke FN, there is an absence of financial capacity (Tables 1 and 6). Internal funding opportunities are limited. As one interviewee noted, the Indian Act (Section 89) is a significant economic barrier in stimulating the local economy and explained that

*“assets on reserves cannot be seized from outside entities ... a bank that finances a company [located on FN land] cannot seize assets, so companies are hesitant to do business with Kahnawà:ke because a bank cannot recover assets”*

(Kahnawà:ke FN, Personal Communication, 25 May, 2010). A healthy local economy would be sustained by business investments to generate taxes, which could then be directed to financing watershed health, source water protection and drinking water management activities.



External funding for infrastructure depends on the availability of federal funds beyond the scope of funds made available for O & M. Although Kahnawà:ke's need for a reservoir was identified as early as 2003 and for a new water line in 2002, there was a lack of funds to support water infrastructure needs until a financial opportunity was provided by EAP (Kahnawà:ke FN, personal communication, 3 September, 2010). Due to the rare opportunity provided by the EAP, Kahnawà:ke's plan to improve their water treatment and reservoir was approved and received \$13 million in funding (INAC, 2009a). Kahnawà:ke's reservoir capacity will be increased significantly to meet health and safety requirements (INAC, 2009a). However, had this rare funding opportunity not presented itself, Kahnawà:ke could not have financed this endeavour.

In terms of internal capacity for the community to pay for or access services, there is an absence of capacity. Kahnawà:ke charges a mandatory \$59 per annum per household for both water and wastewater services. However this fee has not changed in 20 years and does not reflect the rise in input costs (Kahnawà:ke FN, Personal Communication, 20 September, 2010). Water management costs are \$1.2 million annually, of which the community of 2,000 residents generates only \$118,000 (Kahnawà:ke FN, Personal Communication, 25 May, 2010).

In summary, financial capacity supports a wide spectrum of aspects related to integrated water management. A major misunderstanding is that FNs acquire most or all of their funds from governments with ease and therefore, should exhibit the presence of financial capacity. However, interviewees revealed the difficulty in acquiring funding, as witnessed in both participating communities. This coincides with the financial burden encountered in communities with high unemployment and the inability to depend on internal financial resources as a prospective source to ensure financial capacity (OAG, 2005; Smith et al., 2006; Swain et al., 2006). As noted previously, Kitigan Zibi and Kahnawà:ke acknowledge financial resources as a key limiting factor to employing a staff person exclusively for watershed matters, to participate in watershed associations, to monitor watershed health and to generate and collect watershed data. The absence of financial capacity impacts actor networks, information management, human resources and technical capacity. This finding is similar to that of Leach and Pelkey's (2001) study of 37 watersheds, which cited adequate funding as the most important factor for successful watershed management. A lack of financial resources hinders the capacity of stakeholders to plan and achieve watershed goals (Litke and Day, 1998).

### *6.6 Institutional capacity*

This capacity encompasses the regulation, legislation, protocols and plans surrounding watershed management. Institutional capacity incorporates appropriate institutional frameworks and policies to support integrated water initiatives (GWP, 2003). Van der Zaag (2003) suggests that since IWRM is based on relationships amongst water users and between water users and the government, it requires good governance.

In the case of the Kitigan Zibi FN, institutional capacity is present (Tables 1, 7). This capacity relates to locally-initiated environmental policies that are flexible and implementable and have cognizance of long-term benefits (20 years or more), despite federal policies critiqued as neither often reviewed, nor changed periodically due to the bureaucratic nature of the federal government (Kitigan Zibi FN, personal communication, 30 April, 2010). Furthermore, Kitigan Zibi has implemented policies to protect drinking water supplies, particularly in determining the location of community landfills and restricting development near water supplies (Kitigan Zibi FN, personal communication, 30 April, 2010). An observation is that locally directed

efforts have benefitted the protection of drinking water and control of land use activities in community well fields, recharge and watershed water supply areas.

In the case of the Kahnawà:ke FN, there is a presence of institutional capacity (Tables 1, 7). Although interviewees indicated the legal framework set out by INAC was complete and clear, they felt federal legislation was not adaptable due to the nature of government bureaucracy. The presence of the existing capacity is due to Kahnawà:ke's ability to implement policies that are responsive to identified environmental priorities. Kahnawà:ke has implemented policies to protect drinking water supplies, particularly in determining the location of on-community landfills and restricting development near water supplies (Kahnawà:ke FN, personal communication, 25 May, 2010). Kahnawà:ke has the ability to adjust regulations within six weeks, indicating a rapid response time for local environmental concerns (Kahnawà:ke FN, personal communication, 25 May, 2010).

Both participant communities indicated a strong presence of institutional capacity. This is largely due to many indicators being focused on local institutional capacity (e.g., FN implementation of by-laws) rather than institutional capacity at a national level (e.g., review and periodic change of federal policies). It is important to note that at the local scale, FNs experience strong institutional capacity when decision-making and control are within their jurisdiction. Based on Day and Cantwell's (1998) case study, governance was identified as being of the greatest significance for FNs involved in the implementation of integrated land and resource planning.

## **7 Conclusions**

Based on this study, both participant FN communities demonstrate an overall partial capacity to practise IWRM. Kitigan Zibi had a presence of 34 out of a total 79 capacity indicators (43% of overall capacity), while Kahnawà:ke had 38 out of a total of 79 capacity indicators (48% of overall capacity). To achieve sustainable, equitable and collaborative integrated water resources partnerships with FNs as key players, it is important that FNs be engaged in capacity development. However, FN capacities to practise IWRM undoubtedly encounter challenges not common to their non-aboriginal Canadian counterparts, particularly in generating financial resources.

The findings of this study suggest that financial resources are necessary to support FN watershed activities as they relate to technical capacity (e.g., to monitor water quality throughout the watershed), human resources capacity (e.g., to employ personnel that have exclusive responsibility for watershed activities, participation, monitoring and collecting data) and information management capacity (e.g., to generate quality water data). Capacity dimensions are interconnected and tend to overlap. As previously noted, in the case of FNs, the presence of financial capacity plays a significant role in contributing to other capacities. However, the presence of each capacity dimension is necessary for a FN community's overall capacity to practise IWRM.

An additional important finding of this study is the necessity of an actor network capacity. The presence of an actor network capacity serves as a precursor for FNs to participate on a collaborative basis with other stakeholders. FN partnerships, cooperation and communication are critical for participation in IWRM in Québec. However, the absence of an actor network capacity, as demonstrated in this study, is connected with the complexity of the socio-political setting in Québec. Although the processes of addressing language,

cultural identity and political tension are not easily resolved and are beyond the scope of this study, alternatives are necessary to ensure the development of FN capacities to practise watershed management and simultaneously address socio-political concerns.

FN-led watershed councils as suggested by Wilson (2004), FN capacity building partners, customary rights education in watershed associations and addressing jurisdictional complexities while recognising aboriginal self-governance could all serve as a means to address language, cultural identity and political tension issues, thus bridging a colonial past and moving forward with progressive and equitable water management systems inclusive of FN perspectives.

## References

- Al-Jayyousi, O. (2000) 'Capacity building for desalination in Jordan: necessary conditions for sustainable water management, *Desalination*', Vol. 1, No. 141, pp.169–179.
- Alfred, G. (1995) *Heeding the Voices of Our Ancestors: Kahnawà:ke Mohawk Politics and the Rise of Native Nationalism in Canada*, Oxford University Press, Toronto.
- Bailie, R., Bronwyn, E.C. and McDonald, E. (2004) 'Water supply and sanitation in remote indigenous communities – priorities for health development', *Australian and New Zealand Journal of Public Health*, Vol. 28, No. 5, pp.409–414.
- Biswas, A. (1996) 'Capacity building for water management: summary and conclusions', *Water Resources Development*, Vol. 12, No. 4, pp.399–405.
- Blokland, M.W., Alaerts, G.J. and Kaspersma, J.M. (2009) *Capacity Development for Improved Water Management*, UNESCO-IHE, Delft, Netherland.
- Brown, R. (2008) 'Local institutional development and organizational change for advancing sustainable urban water futures', *Environmental Management*, Vol. 41, p.221–233.
- Cap-Net (2006) *Capacity Building for Integrated Water Resources Management*, <http://www.cap-net.org/node/10> (Accessed: October 2009).
- Cap-Net (2008) *Indicators: Implementing Integrated Water Resource Management at River Basin Level*, Cap-Net, Pretoria, South Africa.
- Carter, N., Kreutzwiiser, R. and De Loë, R. (2005) 'Closing the circle: linking land use planning and water management at the local level', *Land Use Policy*, Vol. 22, pp.115–127.
- Chiefs of Ontario (2001) *Drinking Water in Ontario First Nation Communities: Present Challenges and Future Directions for On-Reserve Water Treatment in the Province of Ontario*, Chiefs of Ontario, Ontario, Canada.
- Clarke, M., Riben, P. and Nowgesic, E. (2002) 'The association of housing density and tuberculosis in Canadian First Nation communities', *International Journal of Epidemiology*, Vol. 31, pp.940–945.
- Craig, D. (2003) *Global Sustainable Development: Human Rights, Environmental Rights and Indigenous Peoples*, (Working Paper), Australian Human Rights Centre, Sydney, Australia.
- Dalton, S. (2006) 'Watershed stewardship in Canada', *Horizons*, Vol. 9, No. 1, pp.64–70.
- Day, J. and Cantwell, M. (1998) 'Citizen initiated river basin planning: the salmon watershed example [salmon river watershed roundtable]', *Environments*, Vol. 25, Nos. 2–3, pp.80–92.
- De Carvalho, S., Carden, K. and Armitage, N. (2008) 'Application of a sustainability index for integrated urban water management in southern African cities: case study comparison – Maputo and Hermanus', *Water SA*, Vol. 35, No. 2, pp.144–151.
- De Loë, R., Di Giantomasso, S. and Kreutzwiiser, R. (2002) 'Local capacity for groundwater protection in Ontario', *Environmental Management*, Vol. 29, No. 2, pp.217–233.
- De Loë, R. and Lukovich, D. (2004) 'Groundwater protection on long Island, New York: a study in management capacity', *Journal of Environmental Planning and Management*, Vol. 47, No. 4, pp.517–539.

- Duncan, L. and Bowden, M. (2009) *A Legal Guide to Aboriginal Drinking Water: A Prairie Province Perspective*, Alberta Law Foundation and the Walter and Duncan Gordon Foundation, Edmonton, Alberta.
- Environmental Finance Center (2005) *Financial Capacity Assessment Indicators: Idaho DWSRF*, Boise State University, Boise, Idaho.
- Farahbakhsh, K., Despins, C. and Leidl, C. (2009) 'Developing capacity for large-scale rainwater harvesting in Canada', *Water Quality Research Journal of Canada*, Vol. 44, No. 1, pp.92–102.
- Forster, S. (1997) 'A Water demand management network for southern Africa', in *Management of Water Demand in Africa and the Middle East: Current Practises and Future Needs*, Brooks, D.B., Rached, E. and Saade, M. (Eds.): Chap. 4, International Development Research Centre (IDRC), Ottawa, Canada.
- Franks, T. (1999) 'Capacity building and institutional development: reflections on water', *Public Administration and Development*, Vol. 19, pp.51–61.
- Global Water Partnership (GWP) (2000) *Integrated Water Resources Management. Technical Advisory Committee (TAC) Background Paper No. 4*, Global Water Partnership, Stockholm, Sweden.
- Global Water Partnership (GWP) (2003) *Effective Water Governance. Technical Advisory Committee (TAC) Background Paper No. 7*, Global Water Partnership, Stockholm, Sweden.
- Global Water Partnership (GWP) (2009) *A Handbook for Integrated Water Resources Management in Basins*, Global Water Partnership and International Network of Basin Organizations, Stockholm, Sweden.
- Gracey, M., Williams, P. and Houston, S. (1997) 'Environmental health conditions in remote and rural aboriginal communities in western Australia', *Australian and New Zealand Journal of Public Health*, Vol. 21, No. 5, pp.511–518.
- Gumbo, B., Forster, L. and Arntzen, J. (2005) 'Capacity building in water demand management as a key component for attaining millennium development goals', *Physics and Chemistry of the Earth*, Vol. 30, pp.984–992.
- Harden, A. and Levaliant, H. (2008) *Boiling Point: Six Community Profiles of the Water Crisis facing First Nations within Canada*, Polaris Institute, Ottawa, Canada.
- Hartvelt, F. and Okun, D.A. (1991) 'Capacity building for water resources management', *Water International*, Vol. 16, pp.176–183.
- Health Canada (2005) *Science and Research: Health and the Environment – Critical Pathways*, <http://www.hc-sc.gc.ca/sr-sr/pubs/hpr-rpms/bull/2002-4-environ/method-eng.php> (Accessed: November 2009).
- Health Canada (2008) *Canadian Drinking Water Health Guidelines Development Process*, <http://www.hc-sc.gc.ca/ewh-semt/pubs/water-eau/development-elaboration/appendix-annexe-b-eng.php> (Accessed: January 2012).
- Hill, C., Furlong, K., Bakker, K. and Cohen, A. (2009) 'Emerging issues in water governance and legislation in the Canadian provinces and territories', *Proceedings from Integrated Watershed Management: Navigating Ontario's Future*, Conservation Ontario, Mississauga, Ontario.
- Hrudey, S. (2008) 'Safe water? depends on where you live!', *Canadian Medical Association*, Vol. 178, No. 8, p.975.
- Hundertmark, W. (2008) 'Building drought management capacity in the Mekong River basin', *Irrigation and Drainage*, Vol. 57, pp.279–287.
- Indian and Northern Affairs Canada (INAC) (2003) *National Assessment of Water and Wastewater Systems*, Indian and Northern Affairs Canada, Ottawa, Ontario.
- Indian and Northern Affairs Canada (INAC) (2004) *Water Quality and First Nation Communities*, Indian and Northern Affairs Canada, Ottawa, Ontario.
- Indian and Northern Affairs Canada (INAC) (2007) *Plan of Action for Drinking Water in First Nations Communities: Progress Report – March 22*, Indian and Northern Affairs Canada, Ottawa, Ontario.
- Indian and Northern Affairs Canada (INAC) (2009a) *Canada's Economic Action Plan Delivers Improved Water Treatment System for Kahnawà:ke First Nation in Québec*, <http://www.ainc-inac.gc.ca/ai/mr/nr/j-a2009/nr000000237-eng.asp> (Accessed: May 2010).

- Indian and Northern Affairs Canada (INAC) (2009b) *Canada's Economic Action Plan Delivers New Water Distribution Network for Kitigan Zibi Anishinabeg First Nation in Québec*, <http://www.ainc-inac.gc.ca/ai/mr/nr/j-a2009/nr000000245-eng.asp> (Accessed: May 2010).
- Indian and Northern Affairs Canada (INAC) (2010a) *Aboriginal Community Profile – Kahnawà:ke First Nation*, <http://www.ainc-inac.gc.ca/ai/scr/qc/aqc/prof/Kahnawake-eng.asp> (Accessed: May 2010).
- Indian and Northern Affairs Canada (INAC) (2010b) *Aboriginal Community Profile – Kitigan Zibi First Nation*, <http://www.ainc-inac.gc.ca/ai/scr/qc/aqc/prof/Kitigan-eng.asp> (Accessed: May 2010).
- Indian and Northern Affairs Canada (INAC) (2010) *Canada's Economic Action Plan – Budget 2009 Highlights: Aboriginal and Northern Investments*, <http://www.ainc-inac.gc.ca/ai/arp/bg09/index-eng.asp#rpts> (Accessed: May 2010).
- Indian and Northern Affairs Canada (INAC) (2010d) *First Nations Water and Wastewater Action Plan Progress Report April 2009 – March 2010*, <http://www.ainc-inac.gc.ca/enr/wtr/pubs/prpf/pad10/pad10-eng.asp> (Accessed: 2010).
- Indian and Northern Affairs Canada (INAC) (2011) *Minister Duncan Celebrates the Completion of the Drinking Water and Wastewater Treatment Systems in the Anishinabeg First Nation of Kitigan Zibi*, <http://www.aadnc-aandc.gc.ca/eng/1314728828866> (Accessed: January 2012).
- Ivey, J., De Loë, R., Kreuzwiser, R. and Ferreyra, C. (2006) 'An institutional perspective on local capacity for source water protection', *Geoforum*, Vol. 37, pp.944–957.
- Kempton, K. (2005) *Bridge Over Troubled Waters: Canadian Law on Aboriginal and Treaty Water Rights and the Great Lakes Annex*, Olthuis Kleer Townshend, Toronto, Ontario.
- Kennedy, K., Simonovic, S., Tajada-Guibert, A., Doria, M.F. and Martin, J.L. (2009) *IWRM Implementation in Basins, Sub-Basins and Aquifers: State of the Art Review*, United Nations World Water Assessment Program, Paris, France.
- Kumar, R. (2005) *Research Methodology: A Step-by-Step Guide for Beginners*, Sage Publications Ltd., London, England.
- Lamoree, G. and Harlin, J. (2002) 'Institutional capacity building within the water resources sector of developing countries', *Water International*, Vol. 27, No. 4, pp.542–549.
- Leach, W. and Pelkey, N. (2001) 'Making watershed partnerships work: a review of the empirical literature', *Journal of Water Resources Planning and Management*, Vol. 127, pp.378–385.
- Lebel, P. (2008) 'The capacity of Montreal Lake, Saskatchewan to provide safe drinking water', (*Unpublished MA thesis*), University of Saskatchewan, Saskatoon, SK.
- Lebel, P. and Reed, M. (2010) 'The capacity of Montreal Lake, Saskatchewan to provide safe drinking water: applying a framework for analysis', *Canadian Water Resources Association*, Vol. 35, No. 3, pp.317–338.
- Litke, S. and Day, J.C. (1998) 'Building local capacity for stewardship and sustainability: the role of community-based watershed management in Chilliwack, British Columbia', *Environments*, Vol. 25, Nos. 2–3, pp.91–109.
- MacIntosh, C. (2009) 'Public health protection and drinking water quality on First Nation reserves: considering the new federal regulatory proposal', *Health Law Review*, Vol. 18, No. 1, pp.5–11.
- Macisaac, D. (1996) Peru. Chap. 8 in *Indigenous Peoples and Poverty in Latin America: An Empirical Analysis*, in Patrinos, H.H. and Psacharopoulos, G., (Eds.): World Bank, Washington, DC, pp.165–204.
- MacLeod, M., Browne, A. and Leipert, B. (1998) 'International perspective: issues for nurses in rural and remote Canada', *The Australian Journal of Rural Health*, Vol. 6, No. 2, pp.72–78.
- McGuire, M., Rubin, B., Agranoff, R. and Richards, C. (1994) 'Building development capacity in nonmetropolitan communities', *Public Administration Review*, Vol. 54, No. 5, pp.426–432.
- Ministère du Développement durable, de l'Environnement et des Parcs (2002a) *Bassin Versants*, <http://www.mddep.gouv.qc.ca/eau/bassinversant/redecoupage/inter.htm> (Accessed: April 2010).



- Ministère du Développement durable, de l'Environnement et des Parcs (2002b) *Québec Water Policy*, <http://www.mddep.gouv.qc.ca/eau/politique/index-en.htm> (Accessed: April 2010).
- Ministère du Développement durable, de l'Environnement et des Parcs (2002c) *Le Puits – Démystifier l'Eau Souterraine*, <http://www.mddep.gouv.qc.ca/eau/souterraines/puits/demystifier.htm#besoins> (Accessed: April 2012).
- Mitchell, B. (2006) 'IWRM in practise: lessons from Canadian experiences', *Journal of Contemporary Water Research and Education*, Vol. 135, No. 1, pp.51–55.
- Mkandawire, T. and Mulwafu, W. (2006) 'An analysis of IWRM capacity needs in Malawi', *Physics and Chemistry of the Earth*, Vol. 31, pp.738–744.
- Morin, A. and Cantin, B. (2009) *Strengthening Integrated Water Resource Management in Canada*, Policy Research Initiative, Ottawa, Canada.
- Morris, T., Boyd, D., Brandes, O., Bruce, J., Hudon, M., Lucas, B., Maas, T., Nowlan, L., Pentland, R. and Phare, M. (2007) *Changing the Flow: A Blueprint for Federal Action on Freshwater*, Toronto, Ontario: The Gordon Water Group of Concerned Scientists and Citizens, <http://www.gordonwatergroup.ca/uploads/File/PDF/ChangingtheFlow.pdf> (Accessed: October 2009).
- Mostert, E., van Beek, E., Bouman, N., Hey, E., Savenije, H. and Thissen, W. (1999) 'River basin management and planning', *Proceedings from the International Workshop on River Basin Management*, UNESCO, The Hague, Netherlands.
- O'Connor, D. (2002) *Part Two: Report of the Walkerton Inquiry: A Strategy for Safe Drinking Water – 2002*, <http://www.attorneygeneral.jus.gov.on.ca/english/about/pubs/walkerton/part2/> (Accessed: October 2009).
- Office of the Auditor General (OAG) (2005) *Drinking Water on Ontario First Nations. 2005 Report of the Commissioner of the Environment and Sustainable Development*, [http://www.oag-bvg.gc.ca/internet/English/parl\\_cesd\\_200509\\_05\\_e\\_14952.html](http://www.oag-bvg.gc.ca/internet/English/parl_cesd_200509_05_e_14952.html) (Accessed: October 2009).
- Pahl-Wostl, C., Sendzimir, J., Jeffrey, P., Aerts, J., Berkamp, G. and Cross, K. (2007) 'Managing change toward adaptive water management through social learning', *Ecology and Society*, Vol. 12, No. 2, p.30.
- Phare, M. (2006) 'Whose water is it?', *Horizons*, Vol. 9, No. 1, pp.16–20.
- Raadgever, G.T., Mostert, E., Kranz, N., Interwies, E. and Timmerman (2008) 'Assessing management regimes in transboundary river basins: do they support adaptive management?', *Ecology and Society*, Vol. 13, No. 1, pp.1–14.
- Ramin, V. (2004) 'The status of integrated water resources management in Canada', in Shrubsole, D. (Ed.): *Canadian Perspectives on Integrated Water Resource Management*, Canadian Water Resources Association, Cambridge, Ontario, Chap. 1, pp.1–32.
- Roy, D., Osborne, B. and Venema, H.D. (2009) *Integrated Water Resources Management in Canada*, International Institute for Sustainable Development, Winnipeg, Manitoba.
- Salée, D. (1995) 'Identities in conflict: the aboriginal question and the politics of recognition in Québec', *Ethnic and Racial Studies*, 18 April, Vol. 18, No. 2, pp.277–314.
- Smith, D.W., Guest, R.K., Svrcek, C.P. and Farahbakhsh, K. (2006) 'Public health evaluations of drinking water systems for First Nation reserves in Alberta, Canada', *Journal of Environmental Engineering and Science*, Vol. 5, pp.S1–S17.
- Statistics Canada (2009) *Definition of Census Metropolitan Area and Non-Census Metropolitan Area*, <http://www.statcan.gc.ca/pub/81-004-x/2009004/def/cma-rmr-eng.htm> (Accessed: January 2012).
- Swain, H., Louttit, S. and Hrudey, S. (2006) *Report of the Expert Panel on Safe Drinking Water for First Nations-Volume 1*, Ottawa, Canada: Minister of Public Works Canada and Government Services Canada, [http://www.sdw-eps.gc.ca/rprt/index\\_e.asp](http://www.sdw-eps.gc.ca/rprt/index_e.asp) (Accessed: October 2009).
- Timmer, D., De Loë, R. and Kreutzwiser, R. (2007) 'Source water protection in the Annapolis Valley, Nova Scotia: lessons for building local capacity', *Land Use Policy*, Vol. 24, pp.187–198.
- United Nations (2009) *State of the World's Indigenous Peoples*, Department of Economic and Social Affairs, New York, United Nations, United Nations Educational, Scientific and Cultural Organization, Paris, France.

- United Nations Development Program (UNDP) (1997) *Governance for Sustainable Human Development: A UNDP Policy Document*, <http://mirror.undp.org/magnet/policy/glossary.htm> (Accessed: July 2010).
- United Nations Water (2009) *A Guide to UN Water*, <http://www.unwater.org/flashindex.html> (Accessed: July 2010).
- Van der Zaag, P. (2003) 'Human resource requirements for implementing integrated water resources management', *Proceedings from the 3rd Zimbabwe National Steering Committee Meeting for the Zambezi River Basin*, Harare, Zimbabwe.
- Van der Zaag, P. (2005) 'Irrelevant buzzword? a capacity building and research agenda for South Africa', *Physics and Chemistry of the Earth*, Vol. 30, Nos. 11–16, pp.867–871.
- Van der Zaag, P. and Savenije, H. (2000) 'Conceptual framework for the management of shared river basins: with special reference to the SADC and EU', *Water Policy*, Vol. 2, pp.9–45.
- Watson, N. (2007) 'Collaborative capital: a key to the successful practise of integrated water resources management', in Warner, J. (Ed.): *Multi-Stakeholder Platform for Integrated Water Resource Management*, Ashgate Publishing Ltd., Hampshire, England, Chap. 3, pp.31–48.
- Wilson, P. (2004) 'First Nations integrated watershed management', in Shrubsole, D. (Ed.): *Canadian perspectives on integrated water resource management*, Canadian Water Resources Association, Cambridge, Ontario, pp. 69–83.
- Yillia, P., Bashir, D. and Donkor, E. (2003) 'Partnership approach in capacity building for IWRM – WA-NET initiative', *Proceedings from the 29th WEDC International Conference*, Abuja, Nigeria.
- ROBVQ (2010) Regroupement des organismes de bassins versants du Quebec, <http://www.robvq.ca>
- Statistics Canada (2010a) <http://www.statcan.gc.ca>

## Notes

<sup>1</sup>Watershed zones are river basins identified by the Government of Québec.

<sup>2</sup>The term aboriginal is used here to refer to a person who identifies with, or is a member of a political or cultural entity comprising persons indigenous in Canada (i.e., FNs, Métis, Inuit and Indian), but may or may not be member of an Indian Band or FN (Statistics Canada, 2010a).

<sup>3</sup>Remote areas are places with a population of less than 1,000 and a density of less than 400 persons per square kilometre (Statistics Canada, 2009).

<sup>4</sup>Health Canada (2008) defines maximum acceptable concentrations as established limits for certain substances that are known or suspected to cause adverse effects on health.

# Exploring Collaborative Adaptive Management of Water Resources

Steve Light<sup>1</sup>, Wietske Medema<sup>1</sup>, Jan Adamowski<sup>1</sup>

<sup>1</sup>Department of Bioresource Engineering, McGill University, Ste Anne de Bellevue, Quebec, Canada

Correspondence: Jan Adamowski, Department of Bioresource Engineering, McGill University, Ste Anne de Bellevue, Quebec, H9X 3V9, Canada. Tel: 514-398-7786 Email: jan.adamowski@mcgill.ca

## Abstract

A host of new and "wicked" problems are plaguing today's water resources and managers. The challenges and obstacles stemming from these problems are multidimensional, cumulative, and unprecedented and speak to the need for continuing to explore new approaches in water resources management and restoration efforts. This new class of interdependent problems is explored in this paper and some recently proposed ideas in collaborative adaptive management (CAM) are further developed to help address these types of "wicked" problems. It is argued that collaborative adaptive management, which combines the concepts of adaptive management and collaborative management, can help address the seemingly intractable technical, environmental and social problems inherent in complex social-ecological systems. Because it is important to highlight the importance of induction and emergent understanding under conditions of complexity, the concept of Ecological Policy Design is revisited as it relates to complex problem solving. Other concepts that are further explored and developed in this article include: project optimization that is based on devising composite solutions rather than attempting to 'divide and conquer' individual subsystems; avoidance of 'instability zones'; ecological and restoration efforts that are more 'future responsive'; and the development of alternative hypotheses worthy of rapid prototyping through collaboration. Sustainable solutions are defined as those capable of governing and preserving the vitality of our waterways and the ecosystems that support them. To this end, we conclude that a more collaborative and adaptive approach to water management must be adopted if these types of solutions are to emerge.

**Keywords:** Adaptive Management; Participation; Water Management; Policy Design

## 1. Introduction

Current paradigms and conventional approaches to water resources management have failed to respond to the wide range of stressors that continue to threaten our socio-ecological systems. From the Columbia River salmon recovery to the Everglades to the San Francisco Bay/Delta, it is clear that our current watershed management practices are inadequate; in fact an alarming number of river-related restoration efforts have outright failed. Many rivers, such as the Missouri River, serve a wide range of economic and environmental purposes: from water supply and flood protection to navigation and hydropower to ecosystem vitality. The diverse and often conflicting uses of such waterways create a host of multifaceted problems that have yet to be properly addressed. Not only are the institutional and agency responses to such problems inadequate, but the majority of waterway infrastructures and designs are as well. This is exemplified in Coastal Louisiana and its flawed approach to flood protection and ecological restoration that totally ignores the dominant influence of navigation in the Mississippi Delta. Unusual climatic conditions in 2005 exposed its piecemeal approach to levee construction and maintenance as well as the fractured federal and state governance system that neglected it. In other parts of the world, from Haiti to Cyprus to Australia, long-term shifts in weather patterns have created water resources conflicts, and these conflicts often bring overarching geo-political consequences, as evidenced by the chaotic state of drought-stricken northern Kenya.

A more nuanced and responsive approach to managing and protecting water resources is crucial in order to address the above described types of problems, as well as those related to negative synergies, rising energy costs, climate change, population increases, and decreasing water availability. Some key lessons have been learned regarding limitations and opportunities of different water management approaches through experimentation over the past few decades with both significant management paradigm shifts as well as massive public works to move, rationalize, conserve, and expand the use of freshwater through space and time (Engle et al., 2011:2):

1. A multitude of conflicting interests and pressures on water resources, decisions about allocation and distribution cannot be sustainable if fragmented by competition across scales and sectors;
2. Management that treats different aspects of water, e.g., hydrological, ecological, and socio-economic, separately, ignores their inherent interdependency, possibly at the expense of long term sustainability.

In this light, it is argued that Collaborative Adaptive Management (CAM) can help water resources managers to better address the influx of new and cumulative problems that arise from the unpredictable and complex nature of water systems. By combining the experiential and experimental dimensions of adaptive management with the vertical and horizontal dimensions of collaborative management, CAM can help managers face the daunting challenges that stem from the complexity, non-reducibility, and unpredictability of socio-ecological systems. CAM acknowledges the interdependent nature of society and ecology, thus increasing the likelihood that healthy, adaptive, and sustainable responses can be developed.

Addressing complex and “wicked” socio-ecological problems such as these with a “divide-and-conquer” approach is counterproductive and ignores the interdependent nature of complex adaptive water systems and the restoration projects needed to protect them. The excesses of power and dominion typically seen within traditional institutions in the water sector do not foster sustainable or inclusive solutions; indeed, the era of bureaucratic administrations serving as the ‘final arbiters’ and master controllers of the system needs to come to an end. Today’s multidimensional problems require multidimensional collaboration - amongst experts, managers, stakeholders, and the public at large. New social and physical realities demand not only a more integrated and adaptive paradigm of management, but also an overall shift toward collaborative problem solving. Adaptive co-management, or collaborative adaptive management (CAM), speaks to this need by adopting a more holistic approach to solving the multidimensional problems inherent within water resources management and restoration. CAM recognizes that for any water management solution to be durable, transparent, and accepted, it must be capable of charting courses where uncertainty is high and stakeholder interests are varied.

Although CAM is a very promising approach to allow for more effective water governance and management by focusing on composite rather than piecemeal solutions and relying on a combination of inductive and deductive reasoning, Engle et al. (2011) stress the importance of reconsidering critical elements of in-

---

tegrated and decentralizing approaches to water management if the aim is to create a synergy by combining such approaches with co-management and AM of water resources. These authors highlight, for example, that “integrated water management regimes with certain elements of centralized command and control may be more adaptive and flexible at the expense of democratization and participation” and that “the implementation of integrated water management approaches, and before that command and control institutions, established mechanisms, structures, and processes may constrain the transition of water management into a new framework that is both integrated and adaptive”.

This paper is aimed to offer some reflections on the subject of managing water related issues and their practical complexities that are primarily based on literature as well as personal experiences and observations of the first author. Similarly, the exploring cases provided in this paper aim to offer the reader narratives of challenges, successes as well as some lessons learned from these practical examples, and are not meant to offer evidence for propositions provided in this paper. In this light, this exploring discussion paper is aimed to provide tacit knowledge and insights that may support to bridge the divide between theory and practice that can so often be observed with the implementation of (new) approaches to water resources management.

## **2. A new class of problems**

Today’s water managers are confronted with a class of problems that is without precedent. To understand how these unprecedented new problems are affecting water managers, one need only look at the situation in Coastal Louisiana. The hurricane-stricken state represents a complex and interdependent problem in need of a collaborative and sustainable solution. Despite the availability of federal and state funds earmarked for restoration projects, Louisiana is losing a football field’s length worth of wetlands every half hour (personal communication, Troy Constance, March, 2007; Coalition to Restore Coastal Louisiana, 2000). Since 2003, the US Army Corps of Engineers, in partnership with the State of Louisiana, have been actively pursuing solutions, initially under the Louisiana Coastal Area (LCA) Study followed by the State’s 2007 Coastal Restoration and Protection Act. However their solutions have not addressed one of the fundamental issues: Mississippi River navigation (Barry, 1997). If a successful outcome is to be negotiated, then the needs of boat operators and navigators (who require stable water levels) must be balanced with those of the wetlands (that require changing amounts of water). The key question remains why, despite detailed analysis by peerless scholars and practitioners (Boesch et al., 2006), is the state still incapable of devising effective restoration solutions?

The Missouri River is the longest river in the United States and often referred to as the “Big Muddy”. According to the nonprofit conservation organization *American Rivers*, it is also the most endangered river in the country. The Missouri River operates within a straight-jacket of multiple and conflicting functions (e.g., water supply, flood protection, navigation, and hydropower), all of which are dictated by the river’s Master Manual. The situation is complicated by the fact that the river flows through nine different states, each with their own set of diverse constituencies. And in the past thirty years, these constituencies have unfortunately been more prone to litigate than collaborate when their interests have conflicted.

In the past three years, however, some recommended steps have been taken to create a river-based, system-level organization of stakeholders that is firmly based on collaborative principles. Efforts targeting wildlife preservation have also been made, as seen by the attempts to apply technical adaptive management tools to the recovery of the pallid sturgeon and the piping plover (Light, 2009; NRC, 2002). Although symbols of progress, these responses still fall desperately short of what is needed to overcome the multidimensional and cumulative challenges being faced.

A look at the many other failed river related efforts in the U.S. and it becomes clear just how urgent the need is for a new approach to managing the nation’s water resources. Examples of these shortcomings abound: the ongoing smelt operations within the San Francisco Bay/Delta; the salmon recovery efforts in the Columbia River; the struggle to save the silvery minnow of the Middle Rio Grande, and the efforts of the Red River of the North to reduce flooding and water quality impacts on Lake Winnipeg. Even the Everglades serves as a reflection of failed water resource management: it was decided to approach the issue of water supply from a ‘master plumbing’ angle rather than advancing a more natural-flow regime, a regime



that is essential in order to recover the Everglades' mosaic of ridges, sloughs, wet prairies, tree islands, and wetland complexes.

Most of the river-related infrastructures and waterways within the United States were built circa 1930-1960 and are in desperate need of redesign and replacement; they have exceeded their economic lives and the challenges listed above underscore this. Moreover, climate change and ongoing shifts in key environmental stressors (e.g. population growth, water availability, oil peaking) will continue to compound existing threats to water-related infrastructure over the next twenty to forty years. Both the yawning oscillations in weather patterns - as evidenced by the 1983-84 El Niño event in South Florida (Light, Gunderson and Holling, 1995) - and the unprecedented records of multi-year droughts in Cyprus (Adamowski and Karapataki, 2010), signal a new physical reality that must be addressed.

Throughout the world, water availability and use, along with population growth and energy costs have become mega-drivers spawning "perfect storms" or negative synergies. Norman Myers (1995), one of the founders of the biodiversity movement, was the first to report on this emergent phenomenon. Kates and Clark (1996) expect the number of surprises (including expressions of negative synergies) to double worldwide by 2025. On both a regional and international level, water resource systems have become increasingly vulnerable to many multi-dimensional surprises, which have allowed them to be swept into the vortex of convergence of multi-dimensional surprises of slow, moderate and rapid time steps at landscape scales, creating these negative synergies that threaten the collapse of social and ecological systems in which they are embedded. Haiti and Yemen are examples while more are in the making even if anticipatory action is begun.

In addition to Haiti and Yemen, Africa is another area struggling to overcome its escalating water problems. Amidst the turbulence that grips most of Africa, Kenya has been one of the few mainstays of democracy and stability on the continent; it is also a country lauded for its historic efforts to preserve biodiversity within its national parks and wildlife preserves. Unfortunately, Kenya is not exempt from feeling the consequences of poor water resource management. The nation continues to suffer through a debilitating drought in its north and northwestern regions - one that has been attributed to long-term shifts in weather patterns. Meanwhile, staggering population growth in recent decades has forced average Kenyans to take extraordinary measures to protect their water supplies. The consequences have been widespread: even shepherds have resorted to carrying automatic assault weapons in order to ensure that their flocks have access to the rapidly diminishing water sources. The Kenyan government has responded in part by relocating rural communities away from crucial urban water supply catchment areas where deforestation and land alteration for food production are reducing annual water yields.

Closer to home, the aftermath of Hurricanes Katrina and Rita revealed many of the same dynamics with water resources playing a major role in the unfolding of a much more perilous situation. The level of destruction caused by Hurricane Katrina could have been reduced had the infrastructures governing Coastal Louisiana's waterways not been in such poor condition. The state's flawed approach toward navigation and flood control was exposed when unusual climatic conditions in 2005 triggered a tragedy far beyond most people's expectations. Indeed the question still remains, "How could this have happened in a developed country such as the United States?" Part of the answer is found in the piecemeal approach to levee construction and maintenance that had evolved in response to the erratic pattern of Congressional appropriations. Coastal Louisiana's natural defenses (e.g. ridges and sloughs that buffered New Orleans) had also slowly been collapsing over decades. The two issues combined made an already-destructive hurricane that much more lethal.

The aftermath of the 2005 storms is still being felt by Louisianans today. Hurricane Katrina peeled back the thin veneer of civilization that had been covering a troubled city. To the casual observer, New Orleans is a national treasure that is globally recognized for its distinct culture, music, art, and Mardi Gras celebrations. Despite this colorful reputation, the community's basic day-to-day needs have been historically abused and neglected. Katrina revealed the structural fissures that had been developing over the years within the city, as evidenced by the deplorable conditions of its hospitals, the neglect of the levee system, and the incapacity of the police, fire, and emergency rescue teams to perform their duties. The world watched as the U.S. National Guard came to help a city that could not help itself. And although Kenya and Coastal Louisiana

---

are separated by more than an ocean, and encompass very different societal orders, similar environmental stressors and negative, self-reinforcing dynamics were - and still are - at work.

These are examples of wicked problems; problems that Ludwig (2001) argues have no solution because current methods of management are incapable of solving them. Part of the struggle is that these problems have no definitive formulation, nor any apparent end point. Such multi-dimensional ills, including the imperative of ecological restoration, have no authoritative expert or answer to which to turn. This paper argues that the traditional concept of water managers and their respective institutions as “controllers of the reins of the system”, or as the “final arbiters” of disputes, is outdated and that attention needs to be shifted to underlying assumptions and conflicts in societal and agency values and missions. The power of domination and its commanding efficiencies of bureaucratic administrations in governing espoused by Max Weber (1904), and the means by which to maintain dominion through deceit, despotism, and political manipulation championed by Niccolò Machiavelli (1998), are the bases by which nation states were pulled out of feudalism.

The time has come to combat these excesses of power and dominion with the emergence of a new social reality (Ehrlich, 2009), one that is marked by post-sovereign governance (Karkkainen, 2000) and the power of collaborative problem solving. This will unleash the force of relationship for the commonwealth (Mesle, 2008) and help civilization come to grips with these wicked problems that give sustenance to C.S. Holling’s (1982) observation that “man appears to have forgotten where he comes from and has lost sense of where he is going.”

### *2.1 To Control or adapt?*

The literature on adaptive management (e.g. Gunderson, Holling and Light, 1995; Lee, 1999; Gunderson and Holling, 2002; Pahl-Wostl, 2007) and collaborative management (e.g. Wondolleck and Yaffee, 2000) describes the type of multidimensional challenges inherent in the management of social-ecological systems such as watersheds. The concepts of adaptive management and collaborative management have recently been converging into the concept of ‘adaptive co-management’ or ‘collaborative adaptive management’ (e.g. Olsson et al., 2004; Armitage et al., 2007; and Huitema et al., 2009). Restoration leaders are discovering that the most viable options are those that help us learn new ways of responding to seemingly intractable technical and social problems (Armitage et al., 2009; Kallis et al., 2009; Huitema et al., 2009). Collaborative Adaptive Management (CAM), links both the experiential and experimental learning dimensions of adaptive management with the vertical and horizontal dimensions of collaborative management, and is aimed at improving our understanding of, and ability to respond to, complex social-ecological systems (Armitage et al., 2009). It is argued that this approach should be more frequently leveraged when devising sustainable solutions to interdependent water resources problems.

Collaborative management emphasizes the sharing of rights, responsibilities, and power between different levels and sectors of government and civil society (Huitema et al., 2009), whereas adaptive management (informed by iterative learning about the social-ecological systems and earlier management successes and failures) increases resilience. Experimentation refers to the ‘probing’ of that system which is being managed and the monitoring the system’s responses to interventions. Scholz and Stiftel (2005) state that CAM involves the “evolution of new governance institutions capable of generating long-term, sustainable policy solutions to wicked problems through coordinated efforts involving previously independent systems of users, knowledge, authorities, and organized interests.”

CAM is not a panacea for all water resources problems involving environmental restoration efforts, and will not be appropriate in all cases (Armitage et al., 2009). However, it can be viewed as a very useful tool in a suite of options to help govern and manage social-ecological systems and restoration efforts that are plagued by uncertainty and complexity. The aim of this paper is to explore the potential of CAM in propelling water resources management and restoration efforts forward in the face of uncertainty and complexity. Again, it is important to note the issue stressed by Engel et al. (2011) that more in-depth research is needed in order to understand the differential abilities of water management regimes to incorporate adaptive approaches to water management due to the diversity of implementation models, as well as the maturity of the water management system in its level of decentralization and integration. Their Brazil case study, in fact, indicates that AM may thrive most when long-term environmental sustainability is prioritized over equity

through the representation of a wider group of stakeholder interests.

### *2.2 Divide and conquer?*

Conventional methods and competencies typically dissect large, complex ecological problems into smaller pieces, while assuming that they can be made more tractable and solvable if a “divide and conquer” strategy is used. Breaking down a problem into its various components and solving each part may seem more practical in theory, but may in fact lead to partial solutions. These partial solutions are problematic because they often lack acceptance by the public and their representatives and fail to meet the multiple and conflicting functions requisite to water management, while ignoring both the high degree of interdependence among the key parts of a restoration or watershed management plan, as well as the feedbacks between the socio-economic and physical components of complex adaptive systems such as watersheds. Optimizing subsystems (e.g., irrigation diversions) or individual functions (e.g., navigation), sub-optimizes the whole. For example, until navigation is fully integrated into restoration and protection in Coastal Louisiana, there will be no durable solutions. Solving the multi-dimensionality of “wholes” at the system-level (Gell-Mann, 2001) requires not only addressing the “parts”, but also a redesign of the requisite functions in ways that service the whole (as composites) that is greater than the sum of its parts.

The sediment of the Mississippi River, for example, is dominated by navigation needs, however the delta is sediment starved. As *Rising Tide* (1998) unassailably documents – sediment management is fundamentally a social-cultural and political problem. In the Kissimmee River Restoration Program (KRRP) that began in 1985, the project team planners ignored the legacy of political boundaries and artifacts of accepted but untested precedents and focused on ecosystem form and function rather than individual elements. The design team developed three holistic (i.e. composite) solutions (Walters, Gunderson and Holling, 1992) that managed to perform multiple functions (i.e. water supply, flood control, navigation, sports fishing and ecological restoration) while simultaneously satisfying a diverse group of constituents. These solutions were embedded in the hydrological determinants of ecological integrity for the lower Kissimmee Basin.-The project team and its counterparts in the federal, state and non-profit sectors were largely successful because there were no rules to follow. No one had ever restored a river before, so problems were uncovered and progress toward fitting solutions were designed in collaboration with the various constituents, and technical and scientific experts.

### *2.3 From pieces to composites*

Solving some or all of the components of complex water resources problems involving ecological restoration is sometimes like “optimizing” the parts. However, this collective of “optimized parts” often fails to realize the intent of ecosystem level restoration actions. Because of the interdependence of landscape dynamics, changes in one-function produces a direct change to that feature and most likely will cause changes to other functions. It was thought that restoration of the Kissimmee River regime would devalue the sports fishing industry, but in fact the opposite happened. The high quality fishing areas along the channel were actually remnants of the old meanders.

CAM is a science based, iterative, and inductive learning approach that considers some or all of the components of a project while also respecting the ways in which the new composite functions may be changed. Adaptive management, also referred to as ‘learning by doing’, is based on ‘a more inductive approach of deriving insights from new information’ where ‘dynamic hypotheses guide reasoning and structured argumentation’ (Pahl-Wostl 2007). Collaborative processes are generally represented as cyclical, and can be characterized as ‘highly iterative and nonlinear’ (Ansell and Gash, 2007). Combining both adaptive and collaborative approaches to managing and solving interdependent water resources problems meets the challenges that external influences and stressors require. It is based on the proposition that responsive, durable and authentic solutions require mending and weaving of strategic relationships (economic, ecological and social) in ways that can be mutually beneficial and reinforcing. In essence, ecological recovery and restoration efforts are able to function as Archimedes’ “levers” that invariably put all the water resource related puzzle pieces (i.e. functions and purpose) into play for eventual reconfiguration.

---

#### *2.4 Familiar patterns from previous experience?*

The business world generally understands the value of an analysis and system that focuses on optimization versus maximization of individual objectives. Various commercial enterprises, such as discount retail stores, rely on optimization to maximize profits; they realize that aggregate profits made on sold merchandise is more important than the profit made on each individual item. A similar approach is used within the construction industry. Known as “unbalanced bidding”, many builders will mark up certain high-volume units in order to achieve greater profits while pricing other items in a way that little or no profit is made. In both scenarios, the full power of making tradeoffs among the component parts in order to achieve an overall “greater good” is evident. CAM makes use of this approach and applies it to water resources management and restoration efforts. It helps leaders arrive at solutions to ecological restoration problems, while also addressing the complex societal problems embedded within these restoration processes. Solving problems in this way has a higher degree of success at tendering a long-term, predictable, and stable solution.

Optimizing the overall enterprise – whether in business or ecological restoration – requires allocation judgments involving competing budgetary claims. Some economists (Stakiv, et al., 2001) believe that ecological benefits can be traded off against competing claims. Yes, there is a “balancing” judgment coupled with the optimization for ecological restoration, but “habitat units” do not allow for robust and viable ecosystems. The design team in the KRRP prepared three composite solutions, which met ecological criteria at differing levels of budget allocations. Cost/benefit analysis for ecological restoration assumes implicitly that “incremental” units can be traded off. Not so: the design team for the KRRP made it known that failure to authorize one of the three composite designs would result in a recommendation of “no action”. However, ecosystem structure and function does not come in “halves” or “quarters” but in “whole” forms, as Nature herself intended.

### **3. How can leaders address obstacles and progress in complex settings?**

Progress<sup>1</sup> is the leader’s end game.<sup>2</sup> Constructive action requires identifying and overcoming those obstacles<sup>3</sup> that may threaten progress. Obstacles, as well as methods to overcome them, are varied and numerous. The most intractable problems are typically fraught with uncertainties, complexities, and have multiple dimensions. Each is addressed below:

#### *3.1 Navigating on uncertain seas*

In most ecological restoration efforts, there are no absolute answers; institutional experts are not always capable of solving all “problems”, especially when differences of opinion are unavoidable and competencies are limited. Although well-intentioned, when tactical applications of solutions are applied in excessive and piecemeal ways, they may do more harm than good. These sorts of applications run the risk of oversimplifying the societal and scientific elements of the problem/issue. Furthermore, it is dangerous to assume that just because a certain method helped solve one problem that it will solve another. Tools must fit the problem and not vice versa. CAM is a method of problem solving that can succeed where other methods fail by looking at each ecological restoration effort uniquely and melding scientific solutions with the dynamic of society’s values. Uncertainty is a natural part of a restoration project; the physical, biological and human aspects of such projects are often unique and generally not well understood.

For navigating on the high seas of uncertainty, CAM takes a science-based approach that structures problems as hypotheses, assumptions, and approaches upon which action is based. CAM pays specific attention to how “answers” are designed so that weaknesses in methods and approaches can be understood, communicated, and if possible, avoided. Uncertainties currently take the form of questions such as: “we don’t know how to do this”, “we don’t understand how this and that affect each other”, or “if we do this, what will happen...will it make things better (progress) or make things worse; fall back to a previous or worse condition?” Uncertainty in this form - encourages scientists to model in search of the “right” approach while forgetting that the nature of some uncertainties cannot be effectively overcome without jeopardizing the support for the effort. Also, some stakeholders rely on this quest to eliminate uncertainty because the status quo may be more comfortable to them. The result may be gridlock. The risks associated with restoration progress may be perceived as harming their interests. CAM cannot eliminate uncertainties but it can be

a way to address them without jeopardizing public support. It may also provide a way forward that sufficiently satisfies stakeholders' fears, thus preventing their concerns from materializing into objections.

### *3.2 Complexity*

Complexity marked by both emergence and multiple causality, is sometimes mixed in with uncertainty. Complexity generally arises when there are many overlapping features whose cause-and-effect relationships are poorly understood. Ecological restoration sciences are already rich in complexity; one must add to this the ever-changing dynamics of the natural, social, and political factors and their context in restoration. Often this dynamic unfolds over many years marked in decades or generations. This three-dimensional nature of complexity (man's application of science, the natural dynamics of the landscape, and the political/social environment) can be faced more successfully if CAM is applied. CAM gives water managers and practitioners a possible means for unraveling the Gordian Knot of ecosystem restoration complexity.

### *3.3 The curse of dimensionality*

The term multidimensional encompasses aspects of both time and space. The most challenging problems encountered in restoration science are a result of this multidimensionality. Temporally, complex problems may manifest themselves into different component parts over different time periods (analogous to "shape-shifting"); they may also change (get better, worse, more complex) with time, or present diminished solution opportunities in the future. Spatially, problems may spread to other areas over time, migrate entirely to other areas, be influenced by adjoining areas, or complicated by differences among geo-political units. Stakeholders often add to the multidimensionality of a problem by introducing diverse views, experiences, missions, visions, and goals. These issues, among others, create special challenges that make the problem more complex, and thus often more difficult to solve. However, applying CAM can help water managers untangle complicated situations and use the aforementioned diversity to their benefit. CAM identifies the multi-dimensional features of a restoration effort and adjusts as changes to these dimensional dynamics warrant. For example, the hydrologic regime of Lake Okeechobee in Florida was dramatically affected by the 1983-84 El Niño event. During this climatic abnormality, the Lake's water levels fluctuated drastically: they reached historic lows followed by historic highs, all the while changing people's perceptions of water management and creating a natural experiment that put water deliveries in the spotlight. Instead of following a contrived schedule of monthly "slugs of water" proposed by Stan Caine (Assistant Secretary for Water and Science, U.S. Department of Interior), based on the results of the natural experiment to improve flows to the Everglades National Park, a rainfall-driven approach was adopted that was subsequently codified in federal statute as the "experimental deliveries program".

### *3.4 Making headway*

Adaptive Management (AM) offers leaders a tool box full of principles, methods and approaches to solve many water resource-related problems:

- Systems Analysis (Holling, 1966)
- Optimization (Walters and Hilborn, 1978)
- Policy Sciences (Clark, Jones and Holling, 1979)
- Decision Analysis (Raiffa, 1970; Hilborn and Walters, 1992)
- Strong Inference (Hilborn and Mangel, 1997)
- Bounded Rationality and Induction (Arthur, 1994)
- Learning and Cognition (Argyris and Schön, 1978; Roling and Jiggins, 2000)
- Dynamic Simulation of multi-dimensional landscape level problems (Walters, Gunderson and Holling, 1992) and marine aquatic systems (Walters et al., 2005)



- 
- Alternative designs, tools and methods for Adaptive Inquiry and resolving uncertainties in forest and watershed problems (Vera Sit and Brenda Taylor, 1998; Anderson et al., 2001)
  - Conceptual Ecological Modeling (Ogden et al., 2005 )

One of the fundamental contributions of AM thinking to strategy formation by Ralph Häfele (1974) is the concept of hypotheticality, which has been overlooked with respect to charting courses where uncertainty is high, risks of policy failure could be calamitous, and adaptive policy design may appropriately be fashioned. The challenge that all adaptive policy designs face is the irreducible risk of taking collective action in the face of the unknown or indeterminacy (i.e., increasing probability and cost of policy failure ( $P_1$ - $P_3$ )). This is a strategic form of adaptive policy analysis drawing from a diversity of case examples that complements the more decision-specific approach (Hilborn and Walters, 1992).

At one end of the policy spectrum are safe-to-fail strategies that maximize resilience, and minimize risk of failure at the potential cost of reward. In the case of the Upper Red Lake walleye fishery the safe-to-fail strategy was ignored, exposing the fisheries to collapse with significant economic loss.<sup>4</sup> The Upper Red Lake walleye fishery that was owned by the Red Lake Band of Chippewa and operated by the Department of Interior since 1917 experienced a collapse in the mid-1990s from over harvesting. The Band assumed leadership in 1997, and with the help of the Minnesota Department of Natural Resources, developed a recovery program. A total of 105 million walleye fry were stocked in the Red Lakes in 1999, 2001, and 2003. These efforts resulted in the three largest walleye year classes in at least 15 years (Rivers, 2006). When the cost and probability of policy failure exceed safe-to-fail, a “learning by doing” strategy may be prudent. Here policy-as-hypothesis – e.g., environmental studies (Schwarz, 2003), natural experiments, active or passive adaptive approaches, or evolutionary decision making (Anderson et al., 2001) – is coupled with monitoring and assessment to chart subsequent policy course corrections with the intent of steering clear of instability zones.

Only decisive action by the Band of Chippewa averted the loss of the Red Lake commercial walleye fishery. Unfortunately, this was not the case for Newfoundland, Canada when it’s once prolific 400-year commercial cod fishery collapsed due to heavy fishing during the 1960-1970s. The precipitous and irretrievable collapse of this world-renowned fishery over the course of 25 years illustrates how quickly instability zones can be entered. The unbridled power of economic forces also increased the probability of policy failure in this case, and managed to transform a once-resilient system into a degraded and broken state. From the collapse of the Newfoundland cod fishery, Walters and Maguire (1996) identified several science-based stock assessment lessons that contributed to policy failure:

- (1) Assessment errors can contribute to overfishing through optimistic long-term forecasts leading to the build-up of overcapacity [of the fishing fleet] or through optimistic assessments which lead to harvest levels being set higher than they should;
- (2) Stock size overestimation is a major risk when commercial catch per effort is used as an abundance trend index, so there is continued need to invest in survey indices of abundance trend no matter what assessment methodology is used; and
- (3) The risk of recruitment overfishing exists and may be high even for very fecund species like cod.

Walters et al. (2005) used Ecosim (a sophisticated modeling platform for exploring the dynamics of marine aquatic systems) to consider changes driven by some factor “outside” the breadth of factors considered in the stock assessment models<sup>5</sup> reflected in Walters and Maguire’s (1996) observations above. After successfully calibrating Ecosim to a number of time-series data sets, they demonstrated that:

“...widespread application of single-species MSY policies would in general cause severe deterioration in ecosystem structure, in particular the loss of top predator species. This supports the long-established practice in fisheries management of protecting at

least some smaller "forage" species specifically for their value in supporting larger piscivores."

This analysis is one example of how policy designs can be made more robust to failure as ecological assessments are enhanced to capture critical aspects of scientific understanding. In the devastating example of the Newfoundland cod fishery, it becomes clear how resilience can be squeezed out of what is supposed to be a safe-to-fail policy. At the opposite end of the spectrum are fail-safe solutions that, by their nature, lack resilience. Häfele (1974) referred to this policy option as "hypotheticality," in which the cost of failure exceeds society's willingness and/or capacity to tolerate. A nuclear physicist by training, Häfele was anticipating eventualities like Chernobyl, Three Mile Island, and groundwater contamination at the Hanford Nuclear Reservation.<sup>6</sup>

Many water resources solutions in the past have been based on "fail-safe" solutions where control and redundancy for safety sake are applied to reduce but never eliminate the potential for catastrophic failure. Humanity's track record for "fail-safe" solutions is mixed at best and has always been conditioned by human error and the unknown (Schrader-Frechette, 1993). With the variability and intensity of weather events extending beyond established periods of record, consulting companies in water resources are beginning to place more attention on the liability associated with certifying a given level of protection associated with storm water or flood control projects. Rediscovering ways to let rivers roam again and non-structural alternatives offer more safe-to-fail alternatives to increasing flood control structures along river banks that render all other upstream and downstream structures less effective.

A country that has been actively rediscovering and inventing new responses to such weather-related phenomena is the Netherlands. In response to rising sea levels, shifts in water storage and runoff patterns from the Alps, and greater implications for flooding, the Dutch have embarked on what is referred to as a "180" degree framework document (Ministry of Transport, Public Works and Water Management 2000). They not only reversed many of the science and engineering traditions for which they are known, but also began changing their 400-year perception of water as the enemy or victim (A Different Approach to Water, 2004). In its simplest description, the Dutch government is literally altering the land to accommodate water and nature. In this way, the most densely populated industrialized nation on the planet is taking steps to increase its resilience to and reduce its reliance on fail-safe solutions. Although this is being done in anticipation of future changes that appear inevitable, the Dutch are also making a statement to the world regarding the need for a new societal response to such human predicaments.

However, as the future departs in radical ways from the past, learning by experience becomes less reliable. The arrows of time and complexity are to a large extent irreversible, leaving us with fewer options than before. And despite all the ambitious efforts of the Dutch to reverse course, there is only so much land that can be turned into water before it threatens to displace their population based on experiences from the past. As the Dutch framework document illustrates, becoming future responsive will require increasing ingenuity, creativity, and shifts in cultural norms if current limits of experience and capacity are to be exceeded. It would appear that the problem of hypotheticality is growing as the risks and costs of policy increase, and that natural capital, not financial capital, has become the limiting factor to further human "growth" and development (Daly, 2005). Therefore, becoming future responsive will require us to gain "experience" by analyzing options in such a way that sophisticated conjectures about the future (Baskerville, 1979) may be developed.

### *3.5 Embracing the counter-intuitive*

The new generation of scientific management ushered in by adaptive management (Walters and Hilborn, 1978; Walters, 1986) relies predominantly on powers of observation (i.e., Bayesian approaches), experimentation (i.e., active and passive adaptive) and deduction (e.g., the powerful "engine of deduction," using dynamic simulation) (Hilborn, Walters and Ludwig, 1995). However, Ludwig (2001) subsequently raised concerns that the extant approaches to "management" fall short of making headway posed by wicked problems. During the workshops at the International Institute of Applied Systems Analysis in Austria leading up to the publishing of *Adaptive Environmental Assessment and Management* (Holling, ed., 1978) inquiries on

---

concerns regarding hypotheticality (Häfele, 1974), the foreclosure of policy options (Walters, 1975), and the question of whether “one can anticipate non-probability events” (Clark and Swain, 1975) were deliberated. When design problems carry a low probability (but high cost) of policy failure, the tactic taken by most nuclear reactor designers is to reduce risks to levels experienced by society as natural hazards (e.g., hurricanes, floods, earthquakes, etc.) based on the assumption that “fail-safe” solutions can never reduce risk to zero (Clark and Swain, 1975).

This notion of anticipating exigencies beyond the realm of experience and non-probabilistic events that immerses decision makers in simulations of hypothetical situations was incorporated as a component of what Clark, Jones and Holling (1979) termed “Ecological Policy Design (EPD)”. Extensive combing of the literature has failed to uncover further development of EPD as a distinct line of inquiry aside from the tradition established by Holling (ed., 1978) and additional research around this topic is warranted.

In today’s world, water resource decision makers are being asked to make more complicated decisions in less time – and with reduced trust on past experiences. Even with the ability to apply the most rigorous of assessments, the long-term consequences of climate change coupled with an increasing number of major environmental stressors (e.g., population growth, oil peaking, mass extinctions, limits to freshwater availability) are creating unique and unprecedented challenges for water managers. However, it can be argued that incorporating certain elements of EPD into a revised definition of CAM will help water managers combat these challenges. The application of both EPD and CAM are important because they help humans overcome the natural deficiencies they have in trying to solve complex problems that involve temporal discontinuities at multiple spatial scales (Gunderson, Holling and Light, 1995; Dörner, 1986, 1996).

To observation, experimentation and deduction, Arthur (1994) adds induction. There are two reasons why deductive reasoning breaks down and those two reasons are what attracted Arthur to the use of induction under conditions of complexity. First, beyond a certain level of complication, our capacity to reason breaks down. This is otherwise known as bounded rationality (Simon, 1956). Second, we cannot depend on the behavior of others to be predictable, so we are forced make conjectures about future behavior. The trap that the skeptic will ceaselessly point out is that this loss of deductive capacity throws reasoning into the land of subjective beliefs, the classic case being the “Black Swan” conundrum. The skeptic will never accept that all swans are white regardless of how many observations he makes that confirm the proposition.

Arthur argues that under conditions of ill-defined problems, humans look for patterns, and simplify them into mental models (a.k.a., schemata or hypotheses) of how the world works from which localized deductions and action are carried out. The results of action either confirm or weaken our beliefs in a given schemata. Mental models are discarded when they cease to perform, and are replaced with ones that will hopefully perform better. The result is inductive “leaps” of understanding, and adaptive behavior (Ashby, 1952, 1956) that helps us improve the efficacy of responses in the face of challenges that differ radically from previous experience. This iteration between induction and deduction in the service of effective action, form the “building blocks” for becoming future responsive.

Arguably, the line of reasoning developed by Clark, Jones and Holling (1979) that attempted to address questions raised by Clark and Swain (1975) has simply been ahead of its time and is in need to be revisited. Advances in understanding complex adaptive systems (Holland, 1998; Gell-Mann, 1994; Kauffman 1995), cognitive learning (Maturana and Varela, 1980, a.k.a., Santiago Theory of Cognition; Röling and Jiggins, 2001) and induction (Arthur, 1994; Kelly, 1996) have affirmed their propositions.

### *3.6 Getting there from here*

The novel ideas captured by Clark et al. (1979) are beyond the scope of this paper. Suffice to say that creating sophisticated conjectures about the future based on “experience” designed in the present (Arthur et al., 2001; Scharmer, 2007), does not displace the deductive logic embedded in the scientific method; they complement each other. Converting “unknowns” into “knowns” requires both inductive and deductive reasoning. The coupling of inductive methods with deductive traditions accelerates the rate at which “unknowns” are converted to “knowns.” This conversion rate yields both the small increments and quantum leaps in understanding, both of which are necessary if policies are to become more future responsive. Inductive reasoning can help fill in gaps in understanding and offer “useable ignorance<sup>7</sup>” (Ravetz, 1986).

By identifying environmental stressors, their potential synergetic effects and influence on policies under formation and their supporting infrastructure can be tested as “hypothetical configurations of relational reality” to increase the robustness of policies under active design.

### 3.7 Why collaboration?

Diversity mirrors complexity and vice versa. Therefore, collaboration offers a diversity of perspectives that can provide water managers with the leverage they need to address multidimensional and complicated problems. Diverse perspectives generate a creative tension that, if properly managed, can create a platform for the emergence of novel solutions that no one expert or authoritative entity could come up with. Additionally, solutions to complex problems require unique and authentic solutions which native knowledge can contribute meaningfully to in terms of policy design. What is the alternative to collaborative risk taking in a world grown increasingly complex, uncertain and menacing? Wicked problems exist in a relational reality that can never be fully penetrated; however, if working in collaboration, diverse interests can create more sophisticated composite solutions than would be otherwise expected. Collaboration is also more likely to generate more robust alternative hypotheses worthy of rapid prototyping. In fact, geneticists, embryologists and theoretical biologists are documenting the evolutionary and developmental power of relationships. Gilbert et al. (2010) summarize recent studies in symbiosis by saying that symbiosis is not the exception, but the rule:

“Kauffman (1995) famously said that ‘*All evolution is coevolution*’. The situation may actually be more intimate. Almost all development may be codevelopment. By codevelopment we refer to the ability of the cells of one species to assist the normal construction of the body of another species.”

From a perspective of evolution, Gilbert et al. (2010) affirms Dawkins’ (1989) proposition that the powers of relation are superior in “selective fitness” to the power of dominion. Not in the sense of fitness in a superior solitary organism, but more “fitting” or appropriate responses in the relational or symbiotic (not parasitic, but in a mutually beneficial way) reality which defines the domain of existence. Moreover, Dawkins applied the concept of neo-Darwinian genetics to human culture. Culture is a reliable conservative replicator that has survival value just as a gene. Genes are reliable conservative replicators that find their expression in phenotypes. The cultural equivalent of a gene, Dawkins argues, is a “meme” that finds its expression in the behavior of humans. Moreover, memes can replicate and adapt much faster than genes. In a penetrating critique of Dawkins’ works, Hayle (1995) prefers to refer to “memes” as “cultural grand narratives”. Hayle views Dawkins’ body of work (e.g., *The Selfish Gene* and *The Blind Watchmaker*) as evidence that the “narrative of individualism” is being eclipsed by evolutionary narratives expressing a relational reality, featuring cooperation over competition.

There is an increasing appreciation (Axelrod, 1984) and expansion in the use of cultural expressions of collaboration, as well as documentation of their potential for achieving “higher order stability” (Dawkins, 1989). Birkhoff (2003) analyzed public participation in water reuse projects in Texas, California, and Georgia and found that better decisions emerged when diverse knowledge, interests, and expertise were part of the decision-making process. When stakeholders were not fully involved in framing, developing, evaluating, and implementing solutions to complex public problems, they tried to find alternative ways to express and meet their interests, which hampered the decision process (Birkhoff, 2003).

### 3.8 Pilot testing, demonstrations, experiments

Another problem solving technique includes the testing of ideas in controlled trials. CAM tests or experiments are meant to include statistically designed evaluations that follow scientific methods. Some phenomena that must be understood to advance progress may defy application of science methods and statistical design. For instance, the domain cannot be controlled well enough to meet science methods, or the test scenarios cannot be pre-determined or created to order. In these cases, rigorous forms of learning by doing,

---

at “safe to fail” scales can yield valuable insights into how things work. This is a way of gaining knowledge that may allow conversion of “unknowns” to “knowns.”

Complex problem solving requires (collaborative) monitoring, data collection, and analysis as a means to develop new knowledge. Sometimes, a structured analytic plan including simulation models during implementation may be required to convert an “unknown” to a “known”. These tests may also satisfy the desire by the public to see incremental and on-the-ground progress towards ecological restoration.

### *3.9 Leading the learning*

The fourth leadership challenge requires monitoring and assessing the incremental changes that result from implementing newly-found knowledge and taking corrective measures, if warranted. Proper monitoring and assessment of the incremental change associated with restoration efforts will equip the leader with validation of the newly found knowledge. In the event some of the newly-found knowledge (understanding) fails to provide the improvements expected, monitoring and assessment would provide early detection and opportunity to re-cycle the new knowledge for fine-tuning and reapplication at a later date. Likewise, validating success in producing expected improvements should be documented and used as a resource for problem solving.

### *3.10 Navigating hidden shoals and whirlpools*

CAM simultaneously manages both conflict and uncertainty by fostering forums of science-based information and collaboration. To establish a foundation for successful CAM, serious efforts must be made to engage the government(s) that have jurisdiction over the respective water resource, and tribal and stakeholder interests must be addressed in order to minimize the likelihood for litigation, gridlock and diminished outcomes. Asymmetries in authority and power in complex problem solving has exacerbated mistrust and contention for decades. Lack of true collaboration not only yields sub-optimal ecological responses to applied scientific knowledge, but also can spell ruin by ignoring the importance of the human side of a restoration endeavor.

With appropriate incentives and thorough trust building, as well as greater support for progress among various interests, agencies will evolve. Collaboration requires more time and patience to develop; however more durable, scientifically sound and widely accepted results will come from the openness, transparency and accountability found in collaboration. At the same time, however, an optimum balance will need to be found to address the potential strain that Engel et al. (2011) discuss between water management theories and their implementation in practice, in particular combining integrated (participatory) and adaptive (flexible) approaches to water management. Although adaptive approaches to water management may appear to be a very suitable answer to dealing with increasing complexities and uncertainties, in reality, it may be challenging to accommodate the intended changes within the framework of previous and current management structures and transitions. In this light, a very recent study on integrated and adaptive governance of water resources in South Africa concludes that the features of integrated and adaptive water governance “comprise a number of synergies and trade-offs, which play out differently in different combinations and contexts” while outlining the need for further empirical research on such trade-offs as well as on “the appropriate degree of features such as participation, flexible institutions and redundancy” (Herrfahrtdt-Pähle, 2012).

## **4. Conclusion**

The new problems facing water managers today are interdependent, complex, and often unpredictable. This new class of problems is exacerbated by climate change, ageing waterway infrastructures, and outdated paradigms of institutional authority and bureaucratic power structures. Natural and man-made water resources systems provide a variety of services, but conventional management methods do not address the complexity and uncertainty surrounding their protection, maintenance, and design.

In this paper, the main features of CAM were discussed and some new concepts explored that may help to reframe CAM in a deeper, more robust, and transdisciplinary perspective, responsive to the unparalleled challenges facing water resources managers and policy makers. CAM, which combines the concepts of



adaptive management and collaborative management, can help address the seemingly intractable technical, environmental and social problems inherent in complex social-ecological systems. It is a hybrid approach to problem solving that allows water managers to respond more effectively to the new physical and social realities of the 21st century. With its 'ecosystem level' view of management and restoration, and belief in policy-as-hypothesis, CAM is an efficient way of managing both the high levels of conflict and uncertainty inherent in water resource management.

This paper presents several flexible and innovative ways to build on CAM by incorporating both old and new lines of cognitive and scientific reasoning. Some of the new concepts discussed in this article include: focusing on composite solutions rather than 'divide and conquer' approaches to problem solving, avoiding instability zones in order to design sustainable policies, developing future responsive policies based on emergent understanding, using collaboration to generate more hypotheses worthy of rapid prototyping, and inclusion of the earlier-proposed but 'neglected' theory of Ecological Policy Design (Clark et al., 1979).

In fleshing out deeper dimensions of CAM and integrating new ones, the aim of this paper is to lengthen the stride in theory and practice as a means of protecting the capacity and integrity of our water resources. When water managers operate under conditions of complexity, deduction is not a sufficient buffer to the influence of bounded rationality, subjectivity, and human conjectures about future behavior. In order to anticipate non-probability events and address wicked problems, rigorous forms of 'learning by doing' at 'safe-to-ail' scales must be incorporated into management. The collapse of the Newfoundland cod fishery not only illustrates the devastating consequences that arise when the probability and cost of policy failure are not properly measured, but also exposes the dangers that arise from entering instability zones.

Despite rigorous assessments, anticipating non-probability events is a difficult task, and the challenge is compounded by the fact that today's water resource managers are being asked to make more complicated decisions in less time with reduced trust on past experience. In developing solutions to these problems, the idea of Ecological Policy Design is revisited as it relates to collective complex problem solving and the human tendency to simplify ill-defined problems into patterns, schemata, or hypotheses. In addition to stressing the power and logic of this largely-overlooked theory, the idea of 'emergent understanding' is also put forth in order to achieve effective action. By using both deduction *and* induction, emergent understanding fosters creative tension between constructing schemata and testing them; it also helps managers avoid instability zones and solve unpredictable and complex problems. When the past does not repeat itself (when new challenges differ radically from previous experience), deduction and induction unite to create inductive 'leaps' of understanding that result in more adaptive and future-responsive behaviors. This new line of reasoning is important because emergent understanding is the foundation upon which future-responsive policies should be built. Indeed, the advancements made in the past twenty years toward understanding complex adaptive systems, cognitive learning, and induction support this proposition, as well as the belief that collaboration is more likely to generate alternative hypotheses worthy of rapid prototyping.

Various global crises and case studies were referenced throughout the paper in order to illustrate and highlight the importance of these propositions. These include the shocking number of failed river-related restoration efforts throughout the United States, the collapse of the oldest, richest fishery on Earth (i.e., the Newfoundland cod fishery), the economic and social conflicts arising in Kenya due to drought conditions, the collapse of social order and loss of life in Coastal Louisiana due to Hurricane Katrina, the rise of unprecedented water crises and aridity levels in Cyprus, and the attempts by the Dutch to alter their land to accommodate rivers in order to reduce flood risk. We argue that adopting a more holistic, integrated, and collaborative approach to water management will provide the necessary protection and flexibility needed to create sustainable solutions to the new class of water resources problems that we face, and will do so in a way that decreases uncertainties without jeopardizing public support and societal values.

An enhanced definition of CAM such as the one proposed in this paper will help leaders develop more robust, long-term, and sustainable solutions to water resources management and restoration efforts. Composite solutions that use emergent understanding to overcome obstacles and deal with uncertainty show a greater depth of understanding and respect for the complex societal problems embedded within water management and policy formation. The ego has been removed from this approach, as it not only acknowledges and respects that uncertainties exist (instead of attempting to eliminate them), but also seeks to ad-

---

dress them strategically so that public and stakeholder support is maintained. Water managers will increase their likelihood of designing more durable and transparent problem solving techniques with this approach.

---

<sup>1</sup> *Progress*: changing the way we think and act, always moving to a higher order of development or evolution.

<sup>2</sup> *Endgame*: represents the final and decisive action. The counter-intuitive is that teaching master's chess, one starts with alternative configurations of the endgame and works backwards to develop patterns of moves that result in successful end results. Design is distinctive from "planning" in this regard. Planning starts with defining the problem and logically progresses to a "preferred alternative."

<sup>3</sup> *Obstacles*: problems, setbacks, challenges, crises, adversity, cascades of adversity...

<sup>4</sup> In 2009, the band harvested about 400,000 pounds of walleye of its quota of 820,000 pounds which it sold to Twin City restaurateurs. From "hook-and-line" commercial fishing the Red Lake band received \$1.75 a pound.

<sup>5</sup> A significant departure from the Hilborn and Walters position (Hilborn and Walters, 1992) which concluded: "We believe the food web modeling approach is hopeless as an aid to formulating management advice; the number of parameters and assumptions required are enormous."

<sup>6</sup> The weapons production reactors were decommissioned at the end of the Cold War, but the manufacturing process left behind 53 million U.S. gallons (204,000 m<sup>3</sup>) of high-level radioactive waste that remains at the site. Washington Department of Ecology. <http://web.archive.org/web/20080624232748/http://www.ecy.wa.gov/features/hanford/hanfordfacts.html>. Retrieved 2010-01-19.

<sup>7</sup> Residual distinctions that are appreciated but not fully understood, which may open new and productive veins of inquiry in the future.

## Acknowledgements

This research was partially funded by the Cyprus Institute and an SSHRC Grant held by Jan Adamowski.

## References

Adamowski, J., and Karapataki, C. (2010). Comparison of multivariate regression and artificial neural networks for peak urban water demand forecasting: the evaluation of different ANN learning algorithms. *Journal of Hydrologic Engineering* 15, 729-743. DOI: 10.1061/(ASCE)HE.1943-5584.0000245

Anderson, J.L., Hilborn, R.W., Lackey, R.T., and Ludwig, D. (2003). Watershed restoration - Adaptive decision making in the face of uncertainty, in R. C. Wissmar & P. A. Bisson (editors), *Strategies for Restoring River Ecosystems: Sources of Variability and Uncertainty in Natural and Managed Systems*. American Fisheries Society. Bethesda, MD.

Armitage et al. (2009). Adaptive co-management for social-ecological complexity. *Frontiers in Ecology and the Environment*, 7(2), 95-102. DOI: 10.1890/070089.

Armitage, D., Berkes, F., and Doubleday, N. (2007). *Adaptive co-management: Collaboration, learning and multilevel governance*. UBC Press, Vancouver, British Columbia, Canada.

Arthur, W., Brian, J.D., Jaworski, J., Jung, M., Nonaka, I., Scharmer, O., and Senge, P. (2000). *Illuminating the Blind Spot: Leadership in the Context of Emerging Worlds*. McKinsey - Society for organizational learning leadership project, Cambridge, MA.

Arthur, W.B. (1994). *Inductive Reasoning and Bounded Rationality*. Paper delivered at the American Economic Association Annual Meeting, Session: Complexity of Economic Theory.

Ashby, R. (1952). *Design for a Brain*. Wiley, Oxford, UK.

Ashby, R. (1956). *An Introduction to Cybernetics*. Chapman and Hall, London, UK.

Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books, New York, USA.

Barry, J.M. (1997). *Rising Tide: The Great Mississippi River Flood and How it Changed America*. Touchstone, NY, USA.

Birkhoff, J. (2003). Community conflict over water reuse. Perspectives from conflict analysis and resolution, in T.W. Hartley (editor), *Water reuse. Understanding public perception and participation*. Water Environment Research Foundation, Alexandria, Virginia, USA.

Boesch, D.F. et al. (2006). January 26 2006 Working group for post-hurricane planning for the Louisiana coast. *A New Framework for Planning the Future of Coastal Louisiana after the Hurricanes of 2005*, Maryland Center for Environmental Science, University of Maryland.

Clark, W.C., Jones, D.D., and Holling, C.S. (1979). Lessons for ecological policy design: A case study of ecosystem management. *Ecological Modelling*, 7(1), 1-53.

Clark, W.C., and Swain, H. (1975). *Hypotheticality, Resilience and Option Foreclosure: Summary Notes of an IIASA Workshop*. IIASA Working Paper.

Coalition to Restore Coastal Louisiana. (2000). *No Time to Lose: Facing the Future of Louisiana and the Crisis of Coastal Land Loss*. Baton Rouge, Louisiana.

Daly, H. (2005). Economics in a full world. *Scientific American*, 293(3), 100-107. DOI: 10.1109/EMR.2005.27010.

Dawkins, R. (1989). *The Selfish Gene*. New edition. Oxford University Press, New York.

Dorner, D. (1986). Assessment of operative intelligence. *Diagnostica*, 32, 290-308.

Ehrlich, P. (2009). Cultural evolution and the human predicament. *Trends in Ecology and Evolution*, 24(8), 410-412. DOI: 10.1016/j.tree.2009.03.015

Engle, N. L., O. R. Johns, M. Lemos, and D. R. Nelson. 2011. Integrated and adaptive management of water resources: tensions, legacies, and the next best thing. *Ecology and Society* 16(1): 19. [online] URL: <http://www.ecologyandsociety.org/vol16/iss1/art19/>

Gell-Mann, M. (1994). *The Quark and the Jaguar: Adventures in the Simple and the Complex*. Little, Brown, and Company, UK.

---

Gilbert, S.F., McDonald, E., Boyle, N., Buttino, N., Gyi, L., Mai, M., Prakash, N., and Robinson, J. (2010). Symbiosis as a source of selectable epigenetic variation: taking the heat for the big guy. *Proc. Royal Soc. Lond*, 365, 371 - 378. DOI: 10.1098/rstb.2009.0245

Gunderson, L. H., and Holling, C.S. (2002). *Panarchy: Understanding Transformations in Human and Natural Systems*. Island Press, Washington, DC.

Gunderson, L. H., Holling, C.S., and Light, S.S. (1995). *Barriers and Bridges to the Renewal of Ecosystems and Institutions*. Columbia University Press, New York.

Häfele, W. (1974). Hypotheticality and the new challenges: The pathfinder role of nuclear energy. *Minerva*, 3, 303–322. DOI: 10.1007/BF01102526.

Hayles, K. (1995). Narratives of evolution and the evolution of narratives, in J.L. Casti and A. Karlqvist, *Cooperation and Conflict in General Evolutionary Processes*, John Wiley and Sons, New York, NY.

Herrfahrdt-Pähle, E. (2012). Integrated and adaptive governance of water resources: the case of South Africa. *Regional Environmental Change*, 2012 – Springer. DOI 10.1007/s10113-012-0322-5.

Hilborn, R., Walters, C.J., and Ludwig, D. (1995). Sustainable exploitation of renewable resources. *Annual Review of Ecology and Systematics*, 26, 45- 67. DOI:10.1146/annurev.es.26.110195.000401

Holling, C.S. (1978). *Adaptive Environmental Assessment and Management*. International Institute for Applied Systems Analysis, Wiley, New York.

Holling, C.S. (1966). The functional response of invertebrate predators to prey density. *Mem. Entomol. Soc. Can.* 48, 1-86.

Holling, C.S. (1980). Forest insects, forest fires and resilience, in *Fire Regimes and Ecosystem Properties*, edited by H. Mooney, J.M. Bonnicksen, N.L. Christensen, J.E. Lotan and W.A. Reiners, USDA Forest Service General Technical Report.

Holling, C.S. (1976). Myths of ecology and energy, in *Future strategies of energy development: A question of scale*. Oak Ridge Associated Universities, Oak Ridge, Tennessee, USA.

Holling, C.S. (1982). Predicting the unpredictable: Is it possible to identify the variable that triggers surprise and change? *UNESCO Courier*, 60-62.

Holling, C.S. (1998). Two cultures of ecology. *Conservation Ecology*, 2(2), 4. URL: <http://www.consecol.org/vol2/iss2/art4/>

Huitema, D., Mostert, E., Egas, W., Moellenkamp, S., Pahl-Wostl, C., and Yalcin, R. (2009). Adaptive water governance: Assessing the institutional prescriptions of adaptive co-management from a governance perspective and defining a research agenda, *Ecology and Society*, 14(1), 26.

Kallis, G., Kiparsky, M., and Norgaard, R. (2009). Collaborative governance and adaptive management: Lessons from California's CALFED water program. *Environmental Science and Policy*, 12, 631-643. DOI: 10.1016/j.envsci.2009.07.002

Kates, R.W., and Clark, W.C. (1996). Expecting the unexpected. *Environment* 38, 28-34.

Karkkainen, B.C., Fung, A. and Sabel, C. (2000). After backyard environmentalism: Toward a performance-based regime of environmental regulation. *American Behavioral Scientist*, 44(4), 690-709.

Kauffman, S. (1995). *At Home in the Universe: The Search for the Laws of Self Organization and Complexity* London. Oxford University Press.

Kelly, K. (1996). *The Logic of Reliable Inquiry*. Oxford University Press, Oxford, UK.

Light, S.S. (2009). Assessment of adaptive management standards of practice pertaining to the Missouri River recovery. *Adaptive Strategies Inc. Report*, St. Paul, MN.

Ludwig, D., Mangel, M., and Haddad, B. (2001). Ecology, conservation, and public policy. *Annu. Rev. Ecol. Syst.*, 32, 481-517.

Maturana, H., and Varela, F.J. (1980). *Autopoiesis and cognition: The organization of the living*. Reidel, Boston.

Mesle, R. (2008). *Process-Relational Philosophy: An Introduction to Alfred North Whitehead*. Templeton Foundation Press, West Conshohocken, PA.

Ministry of Transportation, Public Works and Water Management. (2000). *A New Approach to Water Management*. Reinders Partners, Pim Reinders, The Hague, Netherlands.

Myers, N. (1995). Environmental unknowns. *Science*, 269, 358-360.

National Research Council (2002). *The Missouri River Ecosystem: Exploring the Prospects for Recovery*. National Academies Press, Washington, DC.

Ogden, J.C., Davis, S.M., Jacobs, K.J., Barnes, T., Fling, H.E. (2005). *The use of conceptual ecological models to guide ecosystem restoration in South Florida*. *Wetlands*, 25(4), 795-809. DOI: 10.1672/0277-5212(2005)025[0795:TUOCEM]2.0.CO;2

Olsson, P., Folke, C., Berkes, F. (2004). Adaptive co-management for building resilience in social-ecological systems. *Environmental Management*, 34(10), 75-90.

Pahl-Wostl, C. (2007). The implications of complexity for integrated resources management. *Environmental Modelling and Software*, 22(5), 561-569. Doi: [10.1016/j.envsoft.2005.12.024](https://doi.org/10.1016/j.envsoft.2005.12.024)

*The Prince*. Edited by P. Bondanella. (1998). Oxford University Press, Oxford, UK.



---

Raiffa H. (1970). *Decision analysis – Introductory lectures on choices under uncertainty*. Addison-Wesley.

Ritchey, T. (2007). *Wicked Problems: Structuring Social Messes with Morphological Analysis*. Swedish Morphological Society. Discussion Paper. November, 2007.

Rittel, H., and Webber, M. (1973). Dilemmas in a general Theory of Planning," *Policy Sciences* 4, Elsevier Scientific Publishing, Amsterdam, pp. 155-159.

Rivers, E. (2006). *Red Lake: Back to the Future*. Minnesota Volunteer Magazine. St Paul, MN.

Röling, N., and Jiggins, J. (2000). Agents in adaptive collaborative management: The logic of collective action, in *Biological Diversity: Balancing Interests Through Adaptive Collaborative Management*, edited by L.E. Buck, C.C. Geisler, J. Schelhas, and E. Wollenberg. CRC Press, Boca Raton, FL.

Scharmer, O.C. (2007). *Theory U: Leading from the Future as It Emerges*. Kindle Edition, NY, USA.

Scholz, J.T., and Stiftel, B. (2005). *Adaptive Governance and Water Conflict: New Institutions for Collaborative Planning*. Resources for the Future, Washington, DC.

Schrader-Frechette, K.S. (1993). *Burying Uncertainty. Risk and the Case against Geological Disposal of Nuclear Waste*. University of California Press, Berkeley, USA.

Schwarz A. (2003). *The Impact of State Capacity on Enforcement of Environmental Policies: The Case of China*. *The Journal of Environment and Development*, 12(10), 50-81. DOI: 10.1177/1070496502250438

Simon, H. (1956). A comparison of game theory and learning theory. *Psychometrika*, 21(3), 267-272.

Stakhiv, E., Cole, R., Scodari, P., and Martin, L. (2001). *White Paper on Improving Environmental Benefits Analysis*. Working Draft. Institute for Water Resources, Alexandria, VA.

U.S. Army Corps of Engineers, (2003). *Louisiana Coastal Area, LA - Ecosystem Restoration: Comprehensive Coastwide Ecosystem Restoration Study*. U.S. Army Corps of Engineers. New Orleans, LA. Available online: [http://www.crcl.org/lca\\_menu.htm](http://www.crcl.org/lca_menu.htm).

Walters C.J., Christensen, V., Martell, S.J. and Kitchell, J.F. (2005). Possible ecosystem impacts of applying MSY policies from single-species assessment. *Journal of Marine Science*, 62(3), 558-568. DOI: 10.1016/j.icesjms.2004.12.005.

Walters, C. J., Gunderson, L., and Holling, C.S. (1992). Experimental policies for water management in the Everglades. *Ecological Applications*, 2, 189-202.

Walters, C.J. and R. Hilborn. 1978. Ecological optimization and adaptive management. *Annual Review of Ecology and Systematics*, 8, 157-188.

Walters, C.J. and Maguire, J.J. (1996). Lessons for stock assessment from the northern cod collapse. *Reviews of Fish Biology and Fisheries*, 6, 125-137.

Weber, M. (1904). *The Protestant Ethic and "The Spirit of Capitalism"*. Translated by P. Baehr and G.C. Wells. Penguin Books, 2002.

Wondolleck, J.M., and Yaffee, L. (2000). *Making Collaboration Work*. Island Press, Washington, DC.

## Using support vector regression to predict direct runoff, base flow and total flow in a mountainous watershed with limited data in Uttarakhand, India

JAN ADAMOWSKI

Department of Bioresource Engineering, Faculty of Agricultural and Environmental Sciences, McGill University

**Abstract:** *Using support vector regression to predict direct runoff, base flow and total flow in a mountainous watershed with limited data in Uttarakhand, India.* In the ecologically sensitive Himalayan region, land transformations and utilization of natural resources have modified water flow patterns. To ascertain future sustainable water supply it is necessary to predict water flow from the watersheds as affected by rainfall and morphological parameters. Although such predictions may be made using available process-based models, in mountainous and hilly areas it is extremely difficult to determine the numerous parameters needed to run such models, thus limiting their applicability. Artificial intelligence (AI) based models are a possible alternative in such circumstances. In this study an AI technique, support vector machines (SVM), was used for modeling the rainfall-runoff relationship from three hilly watersheds in the state of Uttarakhand, India. Different SVM models were developed to predict direct runoff, base flow, and total flow based on the daily rainfall, runoff, and morphological parameters collected from each watershed. The results confirm the potential of SVM models in the prediction of runoff, base flow, and total flow in hilly areas.

*Key words:* support vector machine, artificial intelligence, runoff, base flow, total flow, Himalayas, watersheds

### INTRODUCTION

In the Himalayan region, agricultural activities are the main source of the popu-

lation's livelihood. While these activities and other land uses have supported the population they have also placed significant pressure on the land, water, and other natural resources creating ecological imbalances. This poses a serious threat to the sustainable development of the region's water resources (Samra et al. 1999). Given this situation, it is necessary to understand the water flow behaviour in this region.

Rainfall-runoff relationships are a complex hydrological phenomenon influenced by temporal and spatial variability in the watershed characteristics, uncertainty in rainfall patterns, and changes in soil cover and morphological parameters (Tokar and Johnson 1999). To simulate the behaviour of this complex system several conceptual process-based models have been developed to mathematically simulate rainfall-runoff processes on a watershed scale. These models are constructed to approximate the general internal sub-process and physical mechanisms that govern the hydrologic cycle, and are based on important hydrological processes. However, it is difficult to translate these processes into mathematical form and in practical situations such models may not be accurate at prediction as they may require

several input parameters that cannot be accurately determined due to spatial and temporal variability. In addition, in hilly regions it is often extremely difficult to collect information due to problematic topography. In such situations, it may be preferable to consider a direct mapping between the readily measurable inputs and outputs by implementing a simpler data driven model with little consideration of the complex processes.

Artificial intelligence (AI) methods such as artificial neural network (ANN) models may be used in runoff prediction without prior knowledge of the actual complex processes involved. A number of researchers have used ANN models for studying rainfall-runoff relationships and found promising results compared to conceptual models (Smith and Eli 1996; LeRoy et al. 1996; Tokar and Johnson 1999; Gautam et al. 2000; Dibike et al. 2001; Jain and Prasad Indurthy 2003; Castellano-Méndez et al. 2004; Nilsson et al. 2005; Adamowski 2008; Adamowski and Sun 2010). Support vector machines (SVM) are another AI method that has recently been employed for solving hydrological problems. Mukherjee et al. (1997) applied SVM regression on chaotic time series and compared the results with those obtained with different prediction methods such as polynomial, rational, local polynomial, multi-quadrics radial basis functions and neural networks. They reported that SVM provided significantly better results as compared to other methods. Dibike et al. (2001) developed both ANN and SVM to predict stream flow discharge at the watershed level using daily rainfall and

evaporation as inputs, reporting a 15% increase in accuracy of runoff estimation with the SVM model over the ANN model. Bray and Han (2004) explored the applicability of the SVM model for flood forecasting, concentrating on determining a suitable model structure and appropriate parameter values for rain-fall-runoff modeling. They addressed the complexity of the SVM optimization with manual based operation of the method and concluded that for appropriate and effective application of SVM in rainfall-runoff modeling more research is needed. Asefa et al. (2005) used SVM models to predict at seasonal and hourly time scales, reporting promising SVM model performance. Behzad et al. (2009) compared SVM against ANN and ANN-GA models and reported the prediction accuracy of SVM to be as good as or better than those models. Wang et al. (2009) developed SVM and three other artificial intelligence models for the same data set, comparing their predictive ability and reporting strong predictive accuracy from the SVM model.

The overall goal of this paper is to investigate the applicability of SVMs in the prediction of runoff, base flow, and total flow for hilly watersheds. Three small watersheds in the Uttaranchal state of India, located in the mid-Himalayan region, were selected and their rainfall and morphological data used to build and test the models. The data for the three watersheds were randomized for training and testing, thus building models that could generalize predictions for similar geographic/climatic watersheds.

## MATERIAL AND METHODS

### Site description and data

The study data for this research were three watersheds located in the hilly terrain of Uttaranchal, India. The Central Soil and Water Conservation Research and Training Institute in Dehradun, Uttaranchal recorded the necessary watershed data. The three watersheds have been denoted as WS1, WS2 and WS3 with areas of 255, 52 and 163 ha, respectively. All three watersheds have varying morphological characteristics with similar steep slopes of 62–66%. WS1 is predominantly mixed forest and scrub, WS3 is mainly a forested watershed, while WS2 consists mainly of agricultural land. More detailed information on these watersheds is given in Table 1. Total flow of all three watersheds was recorded at the outlet with runoff and base flow computed from total flow. Data

were recorded on a daily basis for the three-year period between July 1, 2001 and June 30, 2004.

### Model description

SVMs are based on Vapnik's statistical learning theory (Vapnik 1995). Let us consider a simple problem where the data set has a linear relationship with  $M$  observations. Each observation consists of a pair: a vector  $x_i \in R^n$ ;  $i = 1, \dots, M$ , and the corresponding response variable  $y_i$ . The final objective of SVM regression is to develop a linear function that can make the best approximation of the dependent response variable. The function can be formulated as follows:

$$y = f(x) = \langle w \cdot x \rangle + b \quad (1)$$

where:

$w, b$  – regression parameters,

$\langle w \cdot x \rangle$  – the dot product of  $w$  and  $x$ .

TABLE 1. General characteristics of the three Sainji watersheds

Category	Watershed characteristics	WS1	WS2	WS3
General features	area [ha]	255	52	163
	length [m]	2,950	1,360	2,100
	relief [m]	1,020	635	870
Shape indicators	circulatory ratio [-]	0.553	0.704	0.705
	compactness coefficient [-]	1.34	1.19	1.18
	elongation ratio [-]	0.610	0.598	0.686
Drainage pattern	drainage density [km/km <sup>2</sup> ]	2.76	3.83	2.2
	time of concentration [min]	14	6.76	9.86
	length of streams [m]	7,050	2,010	3,595
	main channel length [m]	2,950	1,360	2,100
Landuse pattern	agriculture [%]	16.55	22.94	14.87
	forest [%]	36.53	0.64	54.01
	scrubs [%]	46.92	76.42	29.12
Hydrologic soil cover complex	weighted curve number	64.99	69.57	62.57



The optimal regression function can be obtained, according to Gunn (1998) and Cristianini and Shawe-Taylor (2000), by minimizing a function,  $\Psi$ , as follows:

minimize

$$\Psi(w, \lambda) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^M (\lambda_i^- + \lambda_i^+) \quad (2)$$

such that  $(w \cdot x_i + b - y_i) + \lambda_i^+ \geq \varepsilon$

where:

$C$  – regularization constant,

$\lambda_i^-$ ,  $\lambda_i^+$  – slack variables that represent the upper and lower constraint on the regression function.

To optimize this function SVM regression uses a loss function that shows the maximum allowed deviation of the predicted values from the measured one. Some of the commonly used loss functions are Quadratic, Laplace, Huber, and  $\varepsilon$ -insensitive (Gunn 1998). Among these, the  $\varepsilon$ -insensitive loss function was proposed by Vapnik (1995) as a robust loss function to reduce sensitivity to the outliers by focusing on optimizing a bound around the regression function. For this study, the  $\varepsilon$ -insensitive loss function was selected. A SVM regression model based on this function calculates the difference between the predicted and the actual values, and if the differences are less than  $\varepsilon$ , the regression function is considered to be acceptable (Schmola and Scholkopf 1998).

Using Lagrangian multipliers, the solution to the optimization problem of equation (2) can be written as follows:

minimize

$$\begin{aligned} & \varepsilon \sum_{i=1}^M (\alpha_i + \alpha_i^*) + \\ & + \frac{1}{2} \sum_{i,j=1}^M (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) (x_i \cdot x_j) + \\ & + \sum_{i=1}^M y_i (\alpha_i - \alpha_i^*) \end{aligned} \quad (3)$$

subject to:

$$\begin{aligned} & \sum_{i=1}^M (\alpha_i - \alpha_i^*) = 0 \\ & \text{and } 0 \leq \alpha_i, \alpha_i^* \leq C \end{aligned}$$

where  $\alpha_i$ ,  $\alpha_i^*$  – the Lagrange multipliers.

To handle non-linear regression cases the data is linearized by mapping it into a higher dimensional space using Lagrange transformations incorporating kernel functions, so that linear regression functions can be applied. The commonly used kernels are the radial basis function (RBF) kernels, sigmoid kernels, and polynomial kernels (Gunn 1998; Chang and Lin 2001). The RBF kernel, most commonly used in SVM approaches, is defined as follows:

$$K(x, y) = e^{-\gamma(x-y)^2} \quad (4)$$

where  $\gamma$  – kernel parameter.

Using the kernel function, the above-mentioned optimization function equation can be rewritten as:

minimize

$$\frac{1}{2} \sum_{i,j=1}^M (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) + \sum_{i=1}^M y_i (\alpha_i - \alpha_i^*) + \varepsilon \sum_{i=1}^M (\alpha_i + \alpha_i^*)$$

subject to

$$\sum_{i=1}^M (\alpha_i - \alpha_i^*) = 0$$

$$\text{and } 0 \leq \alpha_i, \alpha_i^* \leq C$$

The solution of this problem will yield  $\alpha_i$  and  $\alpha_i^*$  for all  $i = 1, 2, \dots, M$ . It should be mentioned that all the training points within the  $\varepsilon$ -sensitive zone will yield  $\alpha_i$  and  $\alpha_i^*$  equal to zero. The type of kernel function to be used is selected by the user. The user also needs to adjust the kernel-specific parameters, the values of parameters  $\gamma$ ,  $C$ , and  $\varepsilon$ . The selection of the optimal values of these parameters determines the success of the SVM approach for a given problem. For more detailed information, readers are referred to Vapnik (1995), Burges (1998), and Cristianini and Shawe-Taylor (2000).

The SVM regression model is trained using a portion of the data set (e.g., 80%) containing the dependent and independent variables. The remaining 20% of the data (unseen data) is used for testing the predictive accuracy or performance of the developed model. The model is run with different sets of values of  $\gamma$ ,  $\varepsilon$ , and  $C$ , using the training data set, and the optimal values are selected by optimizing the cross-validation error using a five-fold cross-validation technique. Next,

the SVM regression model is built based on these optimum values. The generalization ability and predictive accuracy of the model is determined by using the test data set.

### Method description

- (5) In this study SVM models were trained using a set of data containing both independent and dependent variables. The data set contained seventeen independent variables (inputs), namely: day of the year, rainfall, antecedent precipitation index (API<sub>5</sub>), watershed area, length of the watershed, relief, circulatory ratio (the ratio of the watershed area to the area of a circle of the same perimeter as that of the watershed), compactness coefficient (a ratio between the watershed perimeter and the circumference of a circle with the same area), elongation ratio (the ratio of the diameter of a circle with the same area as the watershed area to the maximum length of the watershed), drainage density (which is a ratio between the total length of the drainage channel to the drainage area of a watershed), time of concentration, length of streams, main channel length, percentage of agricultural area, percentage of forested area, percentage of scrubs area, and runoff calculated with the curve number method. The dependent variables (outputs) were runoff, base flow and total flow. Each output was modeled separately as the SVM structure allows modeling of one output at a time.

The data collected over the three watersheds were collated and randomized, generating data sets that did not represent a single watershed but rather the characteristics of all three. This allowed

development of a generalizing model, that if proven accurate would allow its use in predicting flow for ungauged watersheds of similar geographical and climatic characteristics where past rainfall/runoff are not available. This generalizing ability was based upon the work of Sharda et. al (2006) in determining the most important watershed features in the watersheds that affect the rainfall-runoff relationships (curve number, rainfall, antecedent moisture condition and day of the year).

The available data for modeling were limited to just three watersheds over a period of three years, which is not sufficient for modeling hydrologic behaviours. However, the available data were sufficient for evaluating the potential of the SVM method in modeling rainfall-runoff in hilly watersheds. For comprehensive model development, data from a greater number of watersheds over a longer period would need to be collected. Because of the small available data set a five-fold cross-validation procedure was applied to check the generalization ability of the model. In this procedure, the data was randomized and divided into five equal parts. The models were trained using four parts of the data (80%), and tested with the remaining "unseen" fifth part (20%) to evaluate model performance. The procedure was repeated for all five possible combinations.

The performance of each model was evaluated by regression analysis of the simulated results over observed data. The intercepts and the slopes of the best-fit regression lines were determined and compared with their ideal values of 0 and 1, respectively. Other statistical parameters such as, root mean square er-

ror (RMSE), mean bias error (MBE) and modeling efficiency (EF), were also used for the comparison of the estimated and measured data. RMSE represents the mean distance between measured and estimated data (Kobayashi and Salam 2000), and is generally very sensitive to extreme values. Mean bias error (MBE) displays various features of the overall deviation between estimated and measured data, and can show the general bias of the model predictions. The optimum value of MBE is zero and it can be either positive or negative. Modeling efficiency can be viewed as an indicator of the overall correspondence between the measured and estimated values, and explains the goodness of fit of the predicted values with observed values. In the case of biased data, if the data is strongly correlated, EF gives more acceptable analysis than  $R^2$ . A negative value of EF shows the model prediction to be very poor.

To further explore the ability of SVM models in predicting runoff, base flow and total flow models were developed using separate data categorized for years. Year-based models used two-year data sets for model development and the third "unseen" year data was used for model testing. This resulted in a total of three pairs of training and testing data sets.

## RESULTS AND DISCUSSION

### **Prediction of runoff, base flow and total flow using the total data set**

#### *Runoff prediction*

The statistical results from the five-fold cross-validation for surface runoff are presented in Table 2a. For both training and testing data sets, correlation coef-

TABLE 2. Summary of the results obtained from the SVM method applied on watershed data with the five-fold cross-validation procedure, for training and testing (randomized data)

Runoff												
Fold	Training						Testing					
	r	Slope	Intercept	RMSE	MBE	EF	r	Slope	Intercept	RMSE	MBE	EF
1	0.907	0.688	0.066	0.644	0.0294	0.800	0.956	0.865	0.019	0.682	-0.011	0.912
2	0.937	0.789	0.034	0.644	0.003	0.870	0.789	0.737	0.060	0.579	0.033	0.601
3	0.935	0.780	0.036	0.650	0.003	0.864	0.863	0.695	0.036	0.551	0.006	0.741
4	0.940	0.837	0.022	0.492	0.002	0.881	0.876	0.574	0.031	1.235	-0.055	0.719
5	0.930	0.770	0.027	0.679	-0.008	0.855	0.960	1.135	0.037	0.332	0.048	0.869
Base flow												
Fold	Training						Testing					
	r	Slope	Intercept	RMSE	MBE	EF	r	Slope	Intercept	RMSE	MBE	EF
1	0.827	0.662	0.442	0.435	-0.020	0.683	0.770	0.635	0.468	0.495	-0.043	0.586
2	0.823	0.654	0.458	0.444	-0.020	0.676	0.743	0.589	0.539	0.498	-0.017	0.599
3	0.822	0.649	0.469	0.443	-0.016	0.674	0.790	0.630	0.522	0.463	0.024	0.623
4	0.812	0.626	0.482	0.445	-0.029	0.656	0.814	0.617	0.495	0.481	-0.047	0.656
5	0.820	0.652	0.462	0.442	-0.018	0.672	0.783	0.554	0.567	0.485	-0.037	0.605
Total flow												
Fold	Training						Testing					
	r	Slope	Intercept	RMSE	MBE	EF	r	Slope	Intercept	RMSE	MBE	EF
1	0.925	0.756	0.414	0.692	0.052	0.843	0.937	0.855	0.231	0.888	-0.005	0.878
2	0.928	0.792	0.254	0.784	-0.064	0.854	0.835	0.779	0.321	0.740	-0.002	0.688
3	0.964	0.923	0.109	0.544	-0.009	0.930	0.822	0.907	0.179	0.856	0.045	0.595
4	0.930	0.839	0.199	0.632	-0.041	0.863	0.923	0.607	0.499	1.239	-0.135	0.779
5	0.915	0.776	0.301	0.842	-0.042	0.833	0.879	1.191	-0.234	0.867	-0.039	0.546

r – correlation coefficient, RMSE – root mean square error [mm], MBE – mean bias error [mm], EF – modelling efficiency.

coefficients between the observed and predicted runoff were consistently high in the five-fold test (training: 0.91 to 0.94, testing: 0.79 to 0.96), highlighting the ability of the SVM method to learn the input-output relationship and predicting runoff. The value of the intercepts were close to the ideal value of 0, however, the slopes were slightly lower than the ideal

value of 1, which showed that the model under-estimated values for the high runoff events. This may be explained by the small number of high runoff events in the training data. There were only 123 runoff events out of 3,288 data, with only five events having runoff higher than 20 mm, although this may be attributed both to the monsoonal nature of the wa-

tershed and difficulty in collecting data. The RMSE values for the training and testing sets were 0.49 to 0.68 mm and 0.33 to 1.24 mm, respectively. The corresponding MBE values were  $-0.01$  to  $0.03$  mm and  $-0.06$  to  $0.05$  mm. These reasonably low values indicate a close agreement between the observed and simulated runoff for both training and testing data sets (Table 2a). The EF values were high (0.80 to 0.88 for training and 0.62 to 0.92 for testing data), which confirm good model performance in runoff prediction.

#### *Base flow prediction*

The statistical results from the five-fold cross-validation for base flows are presented in Table 2b. The correlation coefficients for both training and testing data sets were consistently high (training: 0.81 to 0.83, testing: 0.74 to 0.81), which implies that the prediction of base flow is quite satisfactory. The low values of RMSE (0.44 to 0.45 mm and 0.46 to 0.50 mm for the training and testing sets, respectively) and MBE ( $-0.03$  to

$-0.02$  mm and  $-0.05$  to  $0.02$  mm) indicate that the observed and simulated values of base flow were very well matched. The slope and the intercept values (0.63 to 0.66 for slope and 0.44 to 0.48 for intercept) of the best-fit lines indicate that the model slightly underestimated the base flow for higher base flow events and overestimated for lower base flow events. The moderate values of EF (0.66 to 0.68 and 0.59 to 0.66 for training and testing sets, respectively) confirm that the SVM model performance was acceptable in estimating base flow.

#### *Total flow prediction*

The statistical results from the five-fold cross-validation for total flows are presented in Table 2c and illustrated in Figures 1 and 2. The high value of the correlation coefficients for both training (0.92 to 0.96) and testing data sets (0.82 to 0.94) indicated that the SVM models consistently predicted total flows very well. The low values of RMSE (training: 0.54 to 0.84 mm, testing: 0.74 to 1.24 mm) and MBE (training:  $-0.06$  to

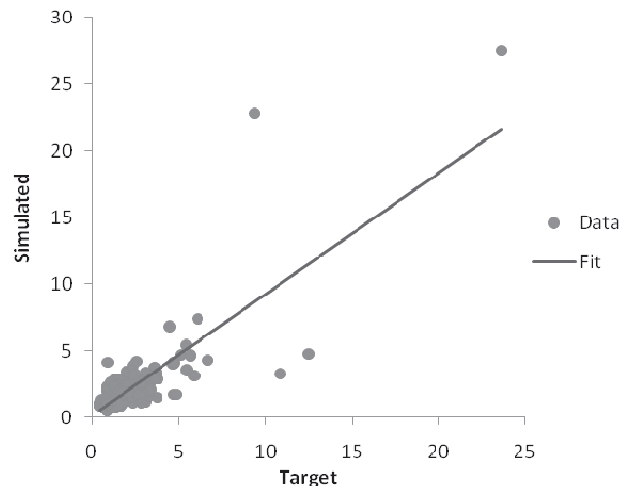


FIGURE 1. Total flow: simulated versus observed values [mm]



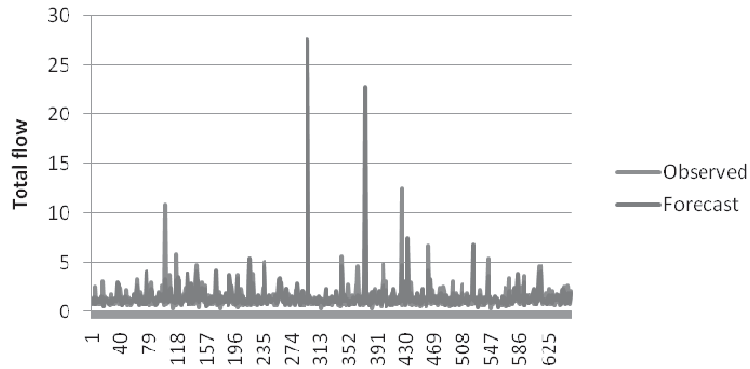


FIGURE 2. Observed and simulated total flow values plotted for each event [mm]

0.05, testing:  $-0.14$  to  $0.04$  mm) along with the high values of EF (training:  $0.83$  to  $0.93$  mm, testing:  $0.55$  to  $0.88$ ) also indicate that the SVM model performance was acceptable in predicting total flows. Figure 1 plots the measured values against the predicted values, with the slope of the line of best fit illustrating the R value. Figure 2 illustrates the measured and predicted values for each flow event.

#### Prediction of runoff, base flow and total flow using year-based data

##### *Runoff prediction*

The statistical results from the model training and validation for runoff prediction using year-based data are reported in Table 3a. The correlation coefficients between the observed and predicted runoff for training data were high (above  $0.93$ ), with the slope and intercept values close to their ideal values for the best fit line. The RMSE, MBE and EF values were closer to their ideal values. Thus, the SVM models were able to develop a very good relationship in training between rainfall-runoff based on the rainfall and

watershed morphological characteristics (Table 3a).

For the test data, the SVM models resulted in high correlation coefficients (above  $0.72$ ), but the slopes of the line of best fit were not very good for all the three possible combinations. The RMSE, MBE and EF values indicated a mismatch between the observed and predicted values of runoff. Such results may be explained by the limited data available in testing, accentuated when the model extrapolates for predictions.

Taking into account the complexity of the procedure of relating rainfall data to runoff, and the limited (three years) data, the results were in the expected range. It is evident that SVM models perform satisfactorily in predicting runoff when the training data set contains extreme runoff values (as was the case with year 1 and 3 for training and year 2 for testing).

##### *Base flow prediction*

The statistical results from the model training and validation for base flow prediction using year-based data are reported in Table 3b. For all three combinations consistently high correlation coefficients (above  $0.80$ ) were obtained

TABLE 3. Summary of the results obtained from the SVM method applied on watershed data with year-based training and testing data

Runoff													
Training						Testing							
Year	r	Slope	Inter-cept	RMSE	MBE	EF	Year	r	Slope	Inter-cept	RMSE	MBE	EF
Y1&Y2	0.978	0.941	0.303	0.421	0.297	0.915	Y3	0.727	0.164	0.324	1.703	0.167	0.271
Y1&Y3	0.930	0.755	-0.009	0.707	-0.043	0.851	Y2	0.903	0.664	0.009	0.554	-0.037	0.787
Y2&Y3	0.951	0.885	-0.046	0.516	-0.064	0.902	Y1	0.928	0.345	-0.012	1.105	-0.072	0.550
Base flow													
Training						Testing							
Year	r	Slope	Inter-cept	RMSE	MBE	EF	Year	r	Slope	Inter-cept	RMSE	MBE	EF
Y1&Y2	0.871	0.733	0.346	0.367	-0.007	0.757	Y3	0.607	0.369	0.801	0.661	-0.134	0.341
Y1&Y3	0.798	0.616	0.513	0.461	-0.017	0.636	Y2	0.794	0.665	0.429	0.479	-0.028	0.628
Y2&Y3	0.838	0.678	0.434	0.439	-0.021	0.700	Y1	0.750	0.619	0.585	0.474	0.098	0.537
Total flow													
Training						Testing							
Year	r	Slope	Inter-cept	RMSE	MBE	EF	Year	r	Slope	Inter-cept	RMSE	MBE	EF
Y1&Y2	0.968	0.889	0.273	0.462	0.114	0.930	Y3	0.659	0.197	1.184	1.880	-0.158	0.299
Y1&Y3	0.935	0.796	0.257	0.757	-0.053	0.866	Y2	0.885	0.782	0.284	0.753	-0.042	0.782
Y2&Y3	0.939	0.863	0.168	0.676	-0.050	0.881	Y1	0.855	0.416	0.865	0.676	-0.050	0.881

r – correlation coefficient, RMSE – root mean square error (mm), MBE – mean bias error (mm), EF – modelling efficiency.

between the observed and predicted base flow for training data. This shows that the SVM models were able to establish a reasonable link between rainfall-runoff events. However, for testing the correlation coefficient values were lower (0.61 to 0.79). Similarly, the slope and intercept values of the best-fit line, RMSE, MBE and EF for the training data sets were consistently closer to their ideal values, whereas this was not the case with the testing data sets.

#### *Total flow prediction*

The statistical results from the model training and validation for total flow prediction using year-based data are reported in Table 3c. The training data sets reported high correlation coefficients (0.94 to 0.97), along with slope and intercepts close to their ideal values. The good RMSE (less than 0.76 mm), MBE (less than 0.11 mm) and EF values (higher than 0.87) also confirmed that the SVM models were able to establish rainfall-total flow relationships in conjunction with the watershed rainfall and morphological characteristics during training. The results for the testing data sets were relatively poor compared to the training data sets; however, the results were satisfactory when the testing data were within the range of the data used for model training.

## CONCLUSIONS

In this study it was found that SVM models have the capability to develop relationships between rainfall and watershed characteristics and runoff, base flow and total flow in hilly watersheds.

The results from the five-fold cross-validations for the training data sets were encouraging, and the results indicate that the models can learn the hidden relationships between the inputs and all the three outputs: runoff, base flow and total flow. The five-fold cross-validations also produced encouraging results when testing the data against “unseen” data.

For the development of SVM models on a yearly basis the training correlations produced were encouraging. However, when tested against “unseen” data, the SVM models did not produce very accurate results. This result may be explained by the smaller data set available for training of these models, and given the encouraging results from training and the five-fold validation tests there is reason to expect that more accurate SVM models for yearly basis may be developed given sufficient training data.

It is concluded that SVM models have the potential to be developed for and used in the accurate prediction of runoff, base flow and total flow in hilly areas. However, it is recommended that further work be undertaken. The utilization of larger data sets to further investigate yearly models in different watersheds needs to be explored. As well, uncertainty related to the forecasting needs to be explored, as does comparing SVM models with other state of the art forecasting methods such as wavelet neural network models.

## Acknowledgements

Financial support for this study was partially provided by an NSERC Discovery Grant (RGPIN 382650-10), as well as a grant provided by the Cyprus Institute, both held by Jan Adamowski.

## REFERENCES

- ADAMOWSKI J. 2008. Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. *Journal of Hydrology* 353, 247–266.
- ADAMOWSKI J., SUN K. 2010: Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology* 390, 85–91.
- ASEFAT T., KEMBLOWSKI M., MCKEE M., KHALIL A. 2005: Multi-time scale stream flow predictions: The support vector machines approach. *Journal of Hydrology* 318: 7–16.
- BRAY M., HAN D. 2004: Identification of support vector machine for runoff modeling. *Journal of Hydroinformatics* 6 (4), 265–280.
- BEHZAD M., ASGHARI K., EAZI M., PALHANG M. 2009: Generalization performance of support vector machines and neural networks in runoff modeling. *Expert Systems with Applications* 36, 7624–7629.
- BURGES C. 1998. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* 2 (2), 121–167.
- CASTELLANO-MÉNDEZ M., GONZÁLEZ-MANTEIGAW., FEBRERO-BANDEM., PRADA-SÁNCHEZ J.M. LOZANO-CALDERÓN R. 2004: Modelling of the monthly and daily behavior of the runoff of the Xallas River using Box-Jenkins and Neural Networks methods. *Journal of Hydrology* 296 (1): 38–58.
- CHANG C., LIN C. 2001: *LIBSVM: a library for support vector machines*. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>. [Last Modification 4.03.2313].
- CRISTIANINI N., SHAWE-TAYLOR J. 2000: *An Introduction to support vector machines and other Kernel-based learning methods*. Cambridge University Press, New York.
- DIBIKE Y.B., VELICKOV S., SOLOMATINE D., ABBOTT M.B. 2001: Model induction with support vector machines. *ASCE Journal of Computing in Civil Engineering* 15 (3), 208–216.
- GAUTAM M.R., WATANABE K., SEAGUSA H. 2000: Runoff analysis in humid forest catchment with Artificial Neural Network. *Journal of Hydrology* 235, 117–136.
- GUNN S. 1998: *Support vector machines for classification and regression*. Technical report, ISIS, Department of Electronics and Computer Science, University of Southampton, Southampton.
- JAIN A., PRASAD INDURTHY S.K.V. 2003: Comparative analysis of event-based rainfall-runoff modeling techniques-deterministic, statistical, and Artificial Neural Networks. *Journal of Hydrologic Engineering* 8 (2), 93–98.
- KOBAYASHI K., SALAM M.U. 2000: Comparing Simulated and measured values using mean squared deviation and its components. *Agronomy Journal* 92, 345–352.
- LeROY P.N., TOKAR A.S., JOHNSON P.A. 1996: Stream hydrological and ecological response to climate change assessed with an Artificial Neural Network. *Limnology and Oceanography* 41 (5), 857–864.
- MUKHERJEE S., OSUNA E., GIROSI F. 1997: *Nonlinear prediction of chaotic time series using support vector machines*. Proceedings of IEEE NNSP VII, 24–26. 09.1997 Amelia Island, 511–520.
- NILSSON P., UVO C., BERNDTSSON R. 2005: Monthly runoff simulation: comparing and combining conceptual and neural network models. *Journal of Hydrology* 321 (4): 344–363.
- SAMRA J.S., DHYANI B.S., SHARMA A.R. 1999: *Problems and prospects of natural resource management in Indian Himalayas – a base paper*. Hill and Mountain Agro-Ecosystem Directorate, NATP. CSWCRTI, 218 Kaulagarh Road, Dehradun.

- SCHMOLA A. SCHOLKOPF A. 1998: *A Tutorial on Support Vector Regression*, NeuroCOLT2 Technical Report NC2-T-R-1998-030.
- SHARDA V.N., PATEL R.M., PRA-SHER S.O., OJASVI P.R., PRAKASH C. 2006: Modeling runoff from middle Himalayan watersheds employing artificial intelligence techniques. *Journal of Agricultural Water Management* 83, 233–242.
- SMITH J., ELI R.N. 1996: Neural-network models of rainfall-runoff process. *Journal of Water Resources Planning and Management* 121 (6), 499–608.
- TOKAR A.S., JOHNSON P.A. 1999: Rain-fall-runoff modeling using Artificial Neural networks. *Journal of Hydrologic Engineering* 4 (3), 232–239.
- Vapnik V. 1995: *The Nature of Statistical Learning Theory*. Springer Verlag, New York.
- WANG W., CHAU K., CHENG C., QIU L. 2009: A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology* 374, 294–306.

**Streszczenie:** Wykorzystanie wektorów wspierających w zależnościach regresyjnych do prognozowania odpływu bezpośredniego i całkowitego w zlewniach górskich przy ograniczonej liczbie danych w zlewni Uttaranchal, Indie. Na ob-

szarach wrażliwych, jakim są Himalaje, zmiany w wykorzystaniu powierzchni obszarów górskich oraz zasobów przyrodniczych modyfikują warunki kształtowania się odpływu. Dla zrównoważonego gospodarowania zasobami wodnymi w tym regionie koniecznym jest prognozowanie odpływu ze zlewni na podstawie opadu i warunków morfologicznych obszaru. Prognozowanie odpływu przy wykorzystaniu modeli deterministycznych jest dosyć trudne i ograniczone ze względu na trudności w identyfikacji wielu parametrów. W pracy zastosowano modele wykorzystujące techniki sztucznej inteligencji (AI) za pomocą wektorów wspierających (SVM) jako alternatywę do modelowania zależności opad-odpływ dla trzech zlewni górskich w stanie Uttaranchal, Indie. Wyniki zawarte w pracy potwierdzają możliwość wykorzystanie metody SVM do prognozowania charakterystycznych wielkości odpływu w warunkach górskich.

*Słowa kluczowe:* metoda wektorów wspierających, sztuczna inteligencja, modelowanie, odpływ, Himalaje, zlewnia góraska

*MS. received 28 June 2013*

**Author's address:**

Department of Bioresource Engineering  
Faculty of Agricultural and Environmental  
Sciences  
McGill University  
21111 Lakeshore Road, Ste-Anne de Bellevue,  
Que., Canada H9X 3V9  
e-mail: jan.adamowski@mcgill.ca





## Towards a Consistent Approach for the Assessment and Redesign of Surface Water Quality Monitoring Networks

Bahaa Khalil<sup>1,2\*</sup> and Jan Adamowski<sup>2</sup>

<sup>1</sup>Department of Civil Engineering, Helwan University, Egypt

<sup>2</sup>Department of Bioresource Engineering, McGill University, Canada

### Introduction

Water resources management in general, and water quality management in particular, requires both knowledge and a full understanding of the processes affecting water quantity and quality. A properly established water quality monitoring program is the only way in which one can understand various processes affecting water quality. It also provides managers and decision makers with the necessary information for effective and sustainable water quality management. Water quality monitoring programs consist of various activities, which include the designation of the monitoring objectives/purposes and the desired information, the design of the monitoring network and the sampling protocol, identification of the necessary laboratory analysis, a plan for data verification and storage, and the design of a data analysis protocol through which the desired information will be obtained. To ensure the utility of the generated data, the sampling processes, sample handling, and storage and laboratory analysis must be performed by professional staff. This is why the design of monitoring networks has generally been the activity that has received the most attention of researchers over the course of the last few decades [1-3]. However, as it stands now, there is no widely accepted established strategy or methodology for designing monitoring networks [2-3]. A logical and consistent design methodology that allows for more efficient and effective data collection and, consequently, more useful outputs, is critically needed. Such an approach would allow not only improved water pollution control recommendations and better allocation of financial resources, but also a better understanding of the ecosystem being monitored [3].

The initial step in the establishment of any monitoring program is to precisely define the monitoring objectives upon which the design of the monitoring network is based. The objectives should precisely indicate why we would do a sample at this specific site, what we want to measure, what is the type of information required, and what is the analysis tool that we intend to use in order to obtain the desired information. Monitoring objectives should be identified in cooperation with different stakeholders such as policy makers, water managers, researchers and the public. A literature review reveals that the most stated water quality monitoring objectives are [3-5] to assess spatial and temporal trends in water quality; to assess compliance with established water quality standards; to assess the effects of natural and anthropogenic factors on the general trends in water quality processes; to determine suitability for various water uses; to assess the effectiveness of water pollution control measures; to assess the general water quality conditions over a wide area; and to facilitate impact assessment studies, mass transport and water quality modeling, as well as other specific research activities. However, it should be emphasized that climate change impacts on surface water quality have not been addressed to date as a water quality monitoring objective [3].

The assessment and redesign of water quality monitoring networks is not a straightforward process, and is based mainly on the monitoring objectives. Therefore, it is very important that the monitoring objectives be carefully and precisely defined. Identification of the monitoring objectives should be specific, clearly stated, and

preferably on a site-by-site basis [3]. The monitoring network design consists of three main aspects: where to measure (sampling sites), what to measure (water quality variables), and how frequently to measure (sampling frequency). Given that water quality is a complex topic, statistical approaches can make a significant contribution regarding the assessment of the performance of a monitoring network [3]. Research into the assessment of the main components of water quality monitoring networks has been ongoing since the 1970s [1,6-16].

In the following sections, statistical approaches that have been used to date for the assessment and redesign of each of the monitoring network components, as well as combined assessments, are summarized. Advantages and disadvantages from the design point of view and recommendations for further work are also presented.

### Sampling Sites

The approaches that have been proposed to date for the assessment of the spatial distribution of the sampling sites have several deficiencies. Some approaches employ only one water-quality variable, even though reliable assessments should be based on a number of water quality variables simultaneously and not sequentially (e.g. stream order approach proposed by Horton [17], Sharp [6,7], Sanders et al. [9]; regression analysis proposed by Tirsch and Male [10], entropy concept proposed by Harmancioglu and Alpaslan [12]). Some approaches are based mainly on the assumption of a linear structure in the data (e.g. regression analysis proposed by Tirsch and Male [10]; entropy concept proposed by Harmancioglu and Alpaslan [12]; multivariate data analysis approaches proposed by Odom [13], Ouyang [14], and Khalil et al. [16]). As well, some of the approaches do not consider reconstruction of information about discontinued sites (e.g. stream order approach proposed by Horton [17]; entropy concept proposed by Harmancioglu and Alpaslan [12]; multivariate data analysis approaches proposed by Odom [13], and Ouyang [14]).

In addition, the approaches that have been proposed to date focus mainly on identifying sampling sites to be discontinued. However, the optimum spatial distribution of sampling sites may involve discontinuing a number of existing sites, while including some of the ungauged sites. This main point was considered by Khalil et al. [16] by incorporating basin characteristics in the assessment approach, and employing cluster analysis and an information index to identify the optimal combination of sites to discontinue, sites to be continuously

\*Corresponding author: Bahaa Khalil, Department of Bioresource Engineering, McGill University, Canada, E-mail: [bahaa\\_khalil@rocketmail.com](mailto:bahaa_khalil@rocketmail.com)

Received December 17, 2012; Accepted December 17, 2012; Published December 24, 2012

Citation: Khalil B, Adamowski J (2013) Towards a Consistent Approach for the Assessment and Redesign of Surface Water Quality Monitoring Networks. *Irrigat Drainage Sys Eng* 2:e113. doi:10.4172/2168-9768.1000e113

Copyright: © 2013 Khalil B, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

measured, and sites to be added. New technologies and techniques need to be explored with respect to the assessment of the spatial distribution of sampling sites. For example, the main advantage of using multivariate data analysis is that it allows for the use of several water quality variables in the assessment procedure simultaneously. However, it relies on the linear structure of the data used. This could be potentially improved through the use of a new generation of nonlinear multivariate data analysis methods (Nonlinear PCA (NLPCA) and Nonlinear Canonical Correlation Analysis (NLCCA) [18]).

## Water Quality Variables

Two main statistical techniques have been proposed to date for the assessment of water quality variables: the correlation regression (CR) approach [1,9] and Principal Component Analysis (PCA) [14]. The advantage of the CR approach is that it allows for the reconstitution of information about discontinued variables. In addition, the modified version of the CR approach, proposed by Khalil et al. [15], overcomes two main deficiencies in the CR approach. Khalil et al. [15] proposed a correlation coefficient threshold above which the variables can be considered highly correlated, and an information index that allows for the identification of the optimal combination of variables to be continuously measured, as well as variables to be discontinued.

These approaches (the CR, its modified version proposed by Khalil et al. [15], and PCA) are mainly based on the assumption of a linear relationship among water quality variables. However, the relationship between physical, biological and chemical variables may be nonlinear. Thus, mutual information and nonlinear regression or artificial neural networks (ANN) may be used instead of linear correlation and regression analysis. Mutual information is a measure of a nonlinear dependence or the amount of redundant information between two variables [19,20]. ANNs are more flexible than regression models in their ability to capture the relationships among water quality variables, and they require less prior knowledge of the system under study [21]. However, ANNs are recommended for the substitution of individual missing values, but not as a record extension technique [22]. ANN as a record extension technique has two main deficiencies: it provides extended records with an underestimated variance, and it is sensitive to the presence of outliers [22]. Thus, a nonlinear record extension technique that can maintain the variance of the extended records, and which is sensitive to the presence of outliers, is required.

## Sampling Frequency

The approaches proposed for the assessment of sampling frequency are based mainly on the monitoring objectives. These approaches address a specific water quality variable at a particular sampling site, and often result in the optimization of sampling frequency based on only one of the monitoring objectives. For instance, Lettenmaier [23] proposed a sampling approach to determine the optimum sampling frequency for trend detection. Sanders and Adrian [8] proposed the confidence interval about the mean as a criterion for the assessment of ambient water quality. Zhou [24] proposed an approach aimed at defining the sampling frequency in terms of periodicity based on harmonic analysis. Mace [25] and Ward et al. [26] described an approach for sampling frequency assessment in order to control the risk of type I and type II errors for compliance with standards assessment. In addition, Tirsch and Male [10] addressed the temporal design of networks using multivariate linear regression, Harmancioglu [11] introduced the entropy concept as a way to determine the optimal sampling intervals, and Khalil et al. [27] proposed the semivariogram to assess sampling frequency based on the range of autocorrelation.

Assessment of sampling frequency by statistical methods often results in the optimization of sampling around one of the monitoring objectives [26]. Water quality monitoring cannot address every information need through one data collection procedure [4]. However, the monitoring network must attempt to meet several objectives simultaneously [3]. Several suggestions to overcome this challenge have been proposed. Whitfield [4] suggested that different sampling frequencies should be used for different monitoring objectives in order to maximize the information gain. Zhou [24] assessed different sampling frequencies for different monitoring objectives, and selected the highest sampling frequency in order to fulfill multiple objectives.

In addition, statistical approaches are used to assess the sampling frequency of a specific water quality variable at a particular sampling site. However, in practice, water quality monitoring networks usually measure several variables at several sites. Ward [28] and Sanders et al. [9] suggested proportional sampling to address this issue. Proportional sampling may be used to distribute a pre-identified total number of samples among sampling sites and/or variables. However, using proportional sampling may provide a number of samples that may not satisfy all the objectives at all the sites for all of the measured variables. Thus, the consequences of applying proportional sampling should be evaluated. The assessment of the sampling frequency for different objectives simultaneously is a possible area for further research.

## Combined Assessment

Although many studies have sought to improve the performance of monitoring networks, most have mainly focused on only one aspect of network design. Few researchers have examined the optimization of different aspects simultaneously [10, 12]. Network assessment and redesign requires combining the assessment of the variables to measure, sampling frequency and sampling sites into one framework. The three main aspects in monitoring design should be assessed simultaneously, and may be linked by a criterion of either cost or information. These assessments can also serve to resolve the trade-off between the number of water quality variables measured and their sampling frequency and sites. Thus, the decision will be either to discontinue more variables in favor of keeping more sampling sites and/or increasing the sampling frequency, or to keep more variables, while decreasing the number of sites and/or the frequency.

## Conclusions

This editorial highlights the main statistical approaches that have been proposed for the assessment and redesign of each of the main components of water quality monitoring networks (sampling sites, water quality variables and sampling frequency). It should be emphasized that monitoring network objective statements should precisely indicate the following important points: why we would do a sample at each site, what we want to measure, what is the type of information required, and what is the analysis tool that we intend to use in order to obtain the desired information? As was mentioned earlier, climate change impacts on surface water quality have not been addressed to date as a water quality monitoring objective; this warrants study in the future.

Recommendations for further research were also highlighted in this editorial for each of the main components of a monitoring network, and can be summarized as follows:

- Assessment and redesign of sampling sites should aim to identify not only the optimal combinations of sites to

discontinue, and sites to be continuously measured, but also sites to be added.

- The use of the recently developed nonlinear multivariate data analyses techniques (e.g. Nonlinear PCA and Nonlinear CCA) should be explored for the assessment and redesign of sampling sites.
- A nonlinear record extension technique that can maintain the variance of the extended records, and which is sensitive to the presence of outliers, is required.
- The assessment of the sampling frequency for different objectives simultaneously is a potentially useful area for further research.
- The consequences of applying proportional sampling for several water quality variables at several sampling sites should be evaluated.
- Simultaneous combined assessment of the three main components of water quality monitoring networks (sampling sites, water quality variables and sampling frequency) using an information or a cost criterion is required.

Finally, the design of a monitoring network needs to be periodically re-assessed and modified accordingly due to shifts in management priorities and/or changing environmental conditions.

#### References

1. Harmancioglu NB, Fistikoglu O, Ozkul SD, Singh VP, Alpaslan MN (1999) Water Quality Monitoring Network Design. Kluwer Academic Publishers, Netherlands.
2. Strobl RO, Robillard PD (2008) Network design for water quality monitoring of surface freshwaters: A review. J Environ Manage 87: 639-648.
3. Khalil B, Ouarda TBMJ (2009) Statistical approaches used to assess and redesign surface water quality monitoring networks. J Environ Monitor 11: 1915-1929.
4. Whitfield PH (1988) Goals and data collection design for water quality monitoring. Water Resources Bulletin. JAWRA 24: 775-780.
5. Harmancioglu NB, Alpaslan MN, Singh VP (1992) Design of water quality monitoring networks: 267-296.
6. Sharp WE (1970) Stream Order as a Measure of Sample Source Uncertainty. Water Resour Res 6: 919-926.
7. Sharp WE (1971) A topologically optimum water sampling plan for rivers and streams. Water Resour Res 7: 1641-1646.
8. Sanders TG, Adrian DD (1978) Sampling Frequency for River Quality Monitoring. Water Resour Res 14: 569-576.
9. Sanders TG, Ward RC, Loftis JC, Steele TD, Adrian DD et al. (1983) Design of Networks for Monitoring Water Quality. Water Resources Publications, USA.
10. Tirsch FS, Male JW (1984) River basin water quality monitoring network

design: options for reaching water quality goals. Proceedings of 20<sup>th</sup> Annual Conference of American Water Resources Associations, AWRA Publications, US.

11. Harmancioglu NB (1984) Entropy concept as used in determination of optimum sampling intervals. Proceedings of International conference on hydraulic engineering software, Portoroz, Yugoslavia.
12. Harmancioglu NB, Alpaslan MN (1992) Water quality monitoring network design: a problem of multi-objective decision making. JAWRA 28: 179-192.
13. Odom KR (2003) Assessment and Redesign of the Synoptic water quality monitoring network in the Great Smoky Mountains National Park. Ph.D. Dissertation, University of Tennessee, Knoxville, USA.
14. Ouyang Y (2005) Evaluation of river water quality monitoring stations by principal component analysis. Water Res 39: 2621-2635.
15. Khalil B, Ouarda TBMJ, St-Hilaire A, Chebana F (2010) A statistical approach for the rationalization of water quality indicators in surface water quality monitoring networks. J Hydrol 386: 173-185.
16. Khalil B, Ouarda TBMJ, St-Hilaire A (2011) A statistical approach for the assessment and redesign of the Nile Delta drainage water-quality-monitoring locations. J Environ Monit 13: 2190-2205.
17. Horton RE (1945) Erosional Development of Streams and their Drainage Basins; Hydrophysical Approach to Quantitative Morphology. Geological Society Am Bull 56: 275-370.
18. Hsieh WW (2004) Nonlinear multivariate and time series analysis by neural network methods. Rev Geophys 42: RG1003
19. Harmancioglu NB (1981) Measuring the information content of hydrological processes by the entropy concept. Journal of civil engineering: 13-38.
20. Harmancioglu N, Yevjevich V (1987) Transfer of hydrologic information among river points. J Hydrol 91: 103-118.
21. ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000) Artificial Neural Networks in Hydrology (II): Hydrologic Applications. J Hydrol Eng 5: 124-137.
22. Khalil B, Adamowski J (2013) Comparison of OLS, ANN, KTRL, KTRL2, RLOC, and MOVE as record-extension techniques for water quality variables. J Hydrol: 13678.
23. Lettenmaier DP (1976) Detection of trends in water quality data from records with dependent observations. Water Resour Res 12: 1037-1046.
24. Zhou Y (1996) Sampling frequency for monitoring the actual state of groundwater systems. J Hydrol 180: 301-318.
25. Mace AE (1964) Sample-size determination. Reinhold, New York, USA.
26. Ward RC, JO Loftis, GB McBride (1990) Design of Water Quality Monitoring systems. Wiley, USA.
27. Khalil B, Abdel-Gawad ST, Abdel-Rashid A, Morsy AM (2004) Sampling frequency assessment for the drainage water quality monitoring in Egypt. Proceedings of the 2<sup>nd</sup> International IWA conference, Automation in Water Quality Monitoring-Automonet, Austria.
28. Ward RC (1978) Evaluating the sampling frequencies of water quality monitoring networks. USEPA, Las Vegas, NV, USA.

Citation: Khalil B, Adamowski J (2013) Towards a Consistent Approach for the Assessment and Redesign of Surface Water Quality Monitoring Networks. Irrigat Drainage Sys Eng 2:e113. doi:10.4172/2168-9768.1000e113

#### Submit your next manuscript and get advantages of OMICS Group submissions

##### Unique features:

User friendly/feasible website-translation of your paper to 50 world's leading languages  
Audio Version of published paper  
Digital articles to share and explore

##### Special features:

250 Open Access Journals  
20,000 editorial team  
21 days rapid review process  
Quality and quick editorial, review and publication processing  
Indexing at PubMed (partial), Scopus, DOAJ, EBSCO, Index Copernicus and Google Scholar etc  
Sharing Option: Social Networking Enabled  
Authors, Reviewers and Editors rewarded with online Scientific Credits  
Better discount for your subsequent articles  
Submit your manuscript at: <http://www.omicsonline.org/submission>





## Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models

Mukesh K. Tiwari<sup>1</sup> and Jan Adamowski<sup>2</sup>

Received 26 July 2012; revised 20 August 2013; accepted 4 September 2013; published 9 October 2013.

[1] A new hybrid wavelet-bootstrap-neural network (WBNN) model is proposed in this study for short term (1, 3, and 5 day; 1 and 2 week; and 1 and 2 month) urban water demand forecasting. The new method was tested using data from the city of Montreal in Canada. The performance of the WBNN method was compared with the autoregressive integrated moving average (ARIMA) and autoregressive integrated moving average model with exogenous input variables (ARIMAX), traditional NNs, wavelet analysis-based NNs (WNN), bootstrap-based NNs (BNN), and a simple naïve persistence index model. The WBNN model was developed as an ensemble of several NNs built using bootstrap resamples of wavelet subtime series instead of raw data sets. The results demonstrated that the hybrid WBNN and WNN models produced significantly more accurate forecasting results than the traditional NN, BNN, ARIMA, and ARIMAX models. It was also found that the WBNN model reduces the uncertainty associated with the forecasts, and the performance of WBNN forecasted confidence bands was found to be more accurate and reliable than BNN forecasted confidence bands. It was found in this study that maximum temperature and total precipitation improved the accuracy of water demand forecasts using wavelet analysis. The performance of WBNN models was also compared for different numbers of bootstrap resamples (i.e., 25, 50, 100, 200, and 500) and it was found that WBNN models produced optimum results with different numbers of bootstrap resamples for different lead time forecasts with considerable variability.

**Citation:** Tiwari, M. K., and J. Adamowski (2013), Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models, *Water Resour. Res.*, 49, 6486–6507, doi:10.1002/wrcr.20517.

### 1. Introduction

[2] Effective and optimized operation and management of urban water resources are critical [Jain and Ormsbee, 2002]. Variation in urban water demand can be attributed to numerous factors such as climatic factors (temperature, rainfall, and humidity) [Altunkaynak *et al.*, 2005; Firat *et al.*, 2009], demographic factors (population, income, people per household, and housing density) [Zhou *et al.*, 2002; Firat *et al.*, 2009], public policy factors (pricing, conservation programs, and education) [Babel *et al.*, 2007; Firat *et al.*, 2009], industrial and commercial factors (nature, size, and productivity) [Zhou *et al.*, 2002], and efficiency and technology [Kayaga and Smout, 2007]. Urban water demand management generally aims to decrease the overall peak demand stemming from the various sources described above.

[3] One of the main purposes of urban water demand forecasting is to match supply with demand at a service level acceptable to consumers [Zhou *et al.*, 2002]. Accurate forecasts allow for optimization of planning, design, management, and operations [Firat *et al.*, 2009], ultimately allowing for efficient allocation of water between competing users and the ecosystem within a river basin [Altunkaynak *et al.*, 2005; Herrera *et al.*, 2010]. Thus, forecasting plays a vital role in socially, economically, and environmentally sustainable water resources management [Caiado, 2010]. Through water demand forecasting, energy use can also be optimized, which is beneficial for both environmental and economic interests [Herrera *et al.*, 2010], especially since energy costs account for 25–30% of total operating costs [Ghiassi *et al.*, 2008]. A distinction between short-term and long-term forecasts is usually made due to their different uses, as well as their differing modeling techniques. Although not specifically defined, short-term forecasts generally include hourly, daily, and weekly forecasts with up to 48 h, 14 day, and 26 week lead times, respectively [Ghiassi *et al.*, 2008]. Long-term forecasts, in comparison, are generally annual and decadal, while monthly forecasts, with up to 24 month lead times, are sometimes classified as medium term. Long-term forecasts can account for economic, demographic, and future climate change variables, which aid in the development, planning, and design of system infrastructure [Jain and Ormsbee, 2002; Ghiassi *et al.*, 2008; Firat *et al.*, 2009; Herrera

<sup>1</sup>Department of Soil and Water Engineering, College of Agricultural and Technology, Anand Agricultural University, Godhra, Gujarat, India.

<sup>2</sup>Department of Bioresource Engineering, McGill University, Ste Anne de Bellevue, Quebec, Canada.

Corresponding author: J. Adamowski, Department of Bioresource Engineering, McGill University, Ste Anne de Bellevue, 21 111 Lakeshore Road, Quebec H9X 3V9, Canada. (jan.adamowski@mcgill.ca)

*et al.*, 2010], as well as in the determination of effective combinations of various water sources to meet water quantity demand and quality standards [Herrera *et al.*, 2010]. Long-term forecasting is also essential for assessing the effectiveness of conservation measures and developing policies and strategies such as water pricing [Babel *et al.*, 2007].

[4] This paper is focused on short-term urban water demand forecasts, which require accurate forecasts of quick, unexpected changes, especially in daily, weekly, and monthly forecasts. Short-term forecasts allow for optimal pump, well, reservoir, and mains operations, balanced allocation amongst urgent water needs, and development of short-term demand management strategies [Jain and Ormsbee, 2002; Kame'enui, 2003; Herrera *et al.*, 2010]. Short-term water demand forecasts also aid in accurate decision making, such as when to implement regulatory water use restrictions in times of water stress or drought [Jain and Ormsbee, 2002; Kame'enui, 2003; Herrera *et al.*, 2010], or when to start drawing from auxiliary supplies [Jain and Ormsbee, 2002]. Urban water supply operators often make the above operational decisions based on experience, but accurate and reliable forecasts can ensure operations are more attuned to demand variability [Zhou *et al.*, 2002]. Important variables in short-term urban water demand forecast modeling include temperature, precipitation, and past water demand data. Water demand exhibits a very complex relationship with all of these input variables, and extracting nonlinearity and nonstationarity from such data is very important. As such, there is a need to develop hybrid models that combine different modeling approaches (that can address nonlinearity, nonstationarity, and uncertainty assessment) to model water demand accurately and reliably. These issues are directly addressed in this paper.

[5] Short-term water demand data generally shows nonlinear and nonstationary behavior [Ghiassi *et al.*, 2008] at multiple spatial and temporal scales [House-Peters and Chang, 2011]. Short-term urban water demand shows diurnal variation, with differing patterns for weekdays, weekends, and holidays, while also showing longer monthly, seasonal, and yearly cycles [Zhou *et al.*, 2002; Caiado, 2010]. In the past three decades, urban water demand modeling has increasingly addressed such behavior, aided by increased data availability and advances in computing methods and power [Zhou *et al.*, 2002; Caiado, 2010]. Short-term demand forecasting has traditionally used linear-regression models, such as multilinear regression and autoregressive integrated moving average (ARIMA) type methods. These methods remain the most common [Adamowski and Chan, 2011], but the inability of linear-regression models to adequately account for nonlinear, nonstationary water demand data has led to the examination of other methods. Zhou *et al.* [2002] and Jain and Ormsbee [2002] applied artificial neural networks (NN) in short-term urban water demand forecasting to address nonlinearity, and subsequent research has generally shown that NNs outperform linear regression techniques in urban water demand forecasting [e.g., Leclerc and Ouarda, 2007; Adamowski, 2008; Sahoo *et al.*, 2009; Adamowski *et al.*, 2012]. Earlier studies demonstrated that NNs have the ability to simulate different water resources time series and to identify the nonlinear relationship between different variables. How-

ever, NN models including other approaches such as ARIMA are generally not able to perform effectively with data that is "noisy" and nonstationary. NN models have shown significant improvement with preprocessed input variables [Cannas *et al.*, 2006; Wu *et al.*, 2010; Tiwari and Chatterjee, 2010b; Adamowski *et al.*, 2012]. Over the course of the last 10 years or so, studies have also explored the potential of wavelet analysis to effectively decompose nonstationary data into sets of new time series at varying scales that can subsequently be used in forecasting models [e.g., Cannas *et al.*, 2006; Adamowski, 2008, 2012]. Adding this data decomposition step prior to feeding the wavelet decomposed data into a NN model has been shown to further improve accuracy, although the use of such hybrid wavelet-neural network (or WNN) models is still very rare in the urban water demand forecasting literature.

[6] Uncertainty in water demand forecasts arising from data, parameters and model structure needs to be quantified in terms of prediction intervals to make forecasts reliable. Data driven models including NN and WNN models are more prone to uncertainty associated with the forecasts [Arhami *et al.*, 2013; Tiwari *et al.*, 2013]. To assess the uncertainty associated with the forecasts obtained using NN and WNN models, the bootstrap technique has been shown to be efficient compared to Bayesian approaches [Hinsbergen *et al.*, 2009] and easy to implement in practice. As such, the bootstrap approach was applied in this study. In earlier studies, ensemble forecasting of water resource variables using the bootstrap technique have increased the reliability of data driven techniques by reducing the variance, and the generated confidence bands using the bootstrap method helped quantify the uncertainty associated with the forecasts [Tiwari and Chatterjee, 2010a, 2011]. For effective and reliable water demand forecasting and management, it is important to know the forecast along with the uncertainty associated with it. Therefore, the bootstrap method was applied in study to assess the uncertainty associated with NN and WNN models by developing BNN and WBNN models, respectively. As discussed, this study coupled the WNN method with bootstrapping, a statistical approach that allows for the quantification of uncertainty through intensive resampling with replacement [Tiwari and Chatterjee, 2010a, 2010b]. Bootstrap-NN (BNN) and wavelet-bootstrap NN (WBNN) models have been one of the newest developments in the literature for discharge and flood forecasting [Sharma and Tiwari, 2009; Tiwari and Chatterjee, 2010a, 2010b, 2011]. While the WBNN method has not yet been applied to urban demand forecasting, preliminary studies by one of the authors of this paper has found the method to be highly accurate and reliable for daily river discharge [Tiwari and Chatterjee, 2011] and hourly flood forecasts [Tiwari and Chatterjee, 2010b].

[7] Several authors have argued that water resources forecasting should explore new hybrid methods that build on the strengths of individual methods, in addition to exploring methods that try to reduce model uncertainty [Jain and Kumar, 2007; Ascough *et al.*, 2008; Srinivasulu and Jain, 2009; Tiwari and Chatterjee, 2009; Maier *et al.*, 2010; Tiwari *et al.*, 2013]. In this study, an attempt is made to incorporate these suggestions by developing hybrid models that generate accurate and reliable urban water demand forecasts and that reduce uncertainty associated with the



forecasts. This study presents, for the first time, the application of the BNN and WBNN method in urban water demand forecasting. The performance of these models is evaluated for different lead times, namely daily, weekly, and monthly forecasting, for the first time. In addition, the effect of the number of bootstrap resamples is assessed for the first time for water demand forecasting, and the selection of an appropriate wavelet transform and the optimum number of decomposition levels for different lead times of water demand forecasting is also explored for the first time.

## 2. Theoretical Background

[8] A theoretical background of wavelet transforms and the bootstrap method is provided since these are both relatively new methods in water resources forecasting. Since the autoregressive integrated moving average (ARIMA) and artificial neural network (ANN) methods are well known, a theoretical background of these two methods is not provided. Adamowski [2008] and Adamowski et al. [2012] provide a detailed background of the ARIMA approach for water resources forecasting, while Bishop [1995] and Haykin [1999] provide further explanations of the general properties of ANNs, and Maier and Dandy [2010] discuss various applications of ANNs in water resources forecasting.

### 2.1. Wavelet Analysis

[9] The basic aim of wavelet analysis is to achieve a complete time-scale representation of localized and transient phenomena occurring at different time scales. Wavelet analysis determines the frequency (or scale) content of a signal and the temporal variation of this frequency content [Heil and Walnut, 1989]. This property of wavelet analysis is different from Fourier analysis that allows for the determination of the frequency content of a signal but fails to determine frequency time dependence. Therefore, the wavelet transform is the tool of choice when signals are characterized by localized high-frequency events or when signals are characterized by a large numbers of scale-variable processes. Because of its localization properties in both time and scale, the wavelet transform allows for tracking the time evolution of processes at different scales in the signal. Water demand time series are characterized by high nonlinearity and nonstationarity, with several meteorological variables and socioeconomic factors characterized by large numbers of scale variable processes affecting it in different ways [Lertpalangsunti et al., 1999; Zhou et al., 2000; Jain and Ormsbee, 2002]. Wavelet analysis can effectively decompose the original time series data into new sets of time series data at varying scales, making nonstationarity more obvious. A brief description of wavelet analysis is described below; for more detailed information on the theory behind wavelets interested readers are directed to refer to Mallat [1998].

[10] The coefficients of the wavelet transform of a continuous-time signal  $f(t)$  are defined by the linear integral operator as

$$W_f(a, b) = |a|^{-\frac{1}{2}} \int_{-\infty}^{+\infty} f(t) \psi * \left( \frac{t-b}{a} \right) dt \quad (1)$$

where  $W_f(a, b)$  is the wavelet coefficient and  $*$  corresponds to the complex conjugate. The wavelet function denoted as  $\psi(t)$ , also known as the mother wavelet can be either real or complex.  $a$  is the scale or frequency factor controlling the dilation ( $a > 1$ ) or contraction ( $a < 1$ ) of the wavelet function, and  $b$  is the time factor affecting the temporal translation of the function.

[11] The mother wavelet function  $\psi(t)$  has finite energy and is mathematically defined as

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0. \quad (2)$$

[12]  $\psi_{a,b}(t)$ , the successive wavelet, can be derived as [Kisi, 2010]

$$\psi_{a,b}(t) = |a|^{-\frac{1}{2}} \psi \left( \frac{t-b}{a} \right) \quad a \in R, \quad a \neq 0, \quad b \in R \quad (3)$$

[13] The wavelet transform searches for correlations between the signal and the wavelet function at different scales of  $a$  and locally around the time of  $b$  to form a contour map known as a scalogram. However, it is inefficient and impractical to compute wavelet coefficients at every resolution level of  $a$  and  $b$ , so values corresponding to the powers of two are often chosen for  $a$  and  $b$ . This arrangement of  $a$  and  $b$  is known as a dyadic grid arrangement and is the simplest and most efficient case for practical purposes. It can be defined as [Mallat, 1989]

$$\psi_{m,n} \left( \frac{t-b}{a} \right) = a_o^{-m/2} \psi * \left( \frac{t-nb_o a_o^m}{a_o^m} \right) \quad (4)$$

where  $m$  and  $n$  are integers that determine the magnitude of wavelet dilation and translation, respectively,  $a_o$  is a specified dilation step greater than 1 (most commonly  $a_o = 2$ ), and  $b_o$  is the location parameter which must be greater than zero (most commonly  $b_o = 1$ ). For a discrete time series  $f(t)$ , assuming  $a_o = 2$  and  $b_o = 1$ , the DWT simplifies as [Kisi, 2010]

$$W_f(m, n) = 2^{-m/2} \sum_{t=0}^{N-1} f(t) \psi * (2^{-m}i - n) \quad (5)$$

where  $W_f(m, n)$  is the wavelet coefficient for the DWT of scale  $a = 2^m$  and location  $b = 2^m n$ .  $f(t)$  is a finite time series ( $t = 0, 1, 2, \dots, N - 1$ ), where the maximum  $t = N$ , defined as an  $M$  integer power of 2 ( $N = 2^M$ ).  $n$  is the time translation parameter in the range  $0 < n < 2^{M-m} - 1$ , and  $m$  is the magnitude dilation parameter with the range  $1 < m < M$ . In this way, a DWT performs a multilevel resolution decomposition of a time series by choosing a discrete scale (integer) for  $m$  and  $n$  to develop a set of wavelet coefficients. For the largest possible wavelet scale such as  $2^m$ , where  $m = M$ , one DWT covers the entire time interval and generates only one coefficient and at the next scale such as  $2^{m-1}$ , two wavelets cover the time interval, generating two coefficients. The same process continues until  $m = 1$  when  $a = 2^1$  (i.e.,  $2^{M-1}$ ), and  $N/2$  coefficients are generated. Finally, the total number of wavelet coefficients produced using

discrete wavelet transformation for a discrete time series of length  $N=2^M$  is  $1+2+4+8+\dots+2M-1=N-1$ . Moreover, the term  $\overline{W}$  can be used to denote the entire signal mean, also called a signal smoothed component. Considering that in wavelet analysis, a time series of length  $N$  is broken into  $N$  components with zero redundancy, the inverse discrete transform can be described as [Nourani *et al.*, 2009]

$$f(t) = \overline{W} + \sum_{m=1}^M \sum_{n=0}^{2^{M-m}-1} W_f(m, n) 2^{-\frac{m}{2}} \psi * (2^{-m}i - n) \quad (6)$$

or further simplified as [Nourani *et al.*, 2009]

$$f(t) = \overline{W}(t) + \sum_{m=1}^M W_m(t) \quad (7)$$

where  $\overline{W}(t)$  is the approximation subsignal at level  $M$ , and  $W_m(t)$  is the detailed subsignal at each level  $m = 1, 2, \dots, M$ .

## 2.2. Bootstrap Technique

[14] The bootstrap technique is a computational, data-driven method that simulates multiple realizations from one data set of a distribution or process. A set of bootstrap samples are created through intensive resampling with replacement after each resampling. This expansion in the number of realizations provides a better understanding of the average and variability of the original, unknown distribution or process, reducing uncertainty [Efron, 1979; Efron and Tibshirani, 1993]. To combine the bootstrap method with NNs in this study, each resampled data set was used to train a single NN. Assume a population of an unknown probability distribution  $F$ , where  $t_i = (x_i, y_i)$  is a realization drawn independently and identically distributed (i.i.d.) from  $F$ ,  $x_i$  is a predictor vector with  $y_i$ , the corresponding output variable, and a  $n$  number of samples are drawn from  $F$ , resulting in a random data set sample, denoted as  $T_n = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ . The empirical distribution function for  $T_n$  is  $\hat{F}$  with a mass of  $1/n$  for each  $t_1, t_2, \dots, t_n$ . Similar to sampling from the unknown distribution  $F$ , an  $n$  number of samples of  $t_i = (x_i, y_i)$  are taken from  $\hat{F}$  that are i.i.d and then replaced. One bootstrap realization would be the resulting random sample set  $T^* = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ , and a set of bootstrap samples of  $T^1, T^2, \dots, T^s, \dots, T^S$  can be produced. The total number of bootstrap samples,  $S$ , usually ranges from 50 to 200 samples [Efron, 1979]. For each  $T^s$ , a NN model is developed and trained using all  $n$  observations. The NN output,  $f_{NN}(x_i, w_s/T^s)$ , is then evaluated using the set  $A$  of observation pairs  $t_i = (x_i, y_i)$  that were not a part of the bootstrap sample  $T^s$ . The performance of the NNs in these validation tests are subsequently averaged, which represents the generalization error for the NN models relative to  $T_n$ . This generalization error is denoted as  $E_0$ , and this can be estimated as [Twomey and Smith, 1998]

$$\hat{E}_0 = \frac{\sum_{s=1}^S \sum_{i \in A} (y_i - f_{NN}(x_i, w_s/T^s))^2}{\sum_{s=1}^S (A)} \quad (8)$$

where  $f_{NN}(x_i, w_s/T^s)$  is, again, the output of the NN developed from the bootstrap sample  $T^s$ , in which  $x_i$  is a particular input vector and  $w_s$  is the weight vector. Finally, the BNN estimate  $\hat{y}(x)$  of all developed NNs is given by the average of the  $S$  bootstrapped estimates [Jia and Culver, 2006]

$$\hat{y}(x) = \frac{1}{S} \sum_{s=1}^S f_{NN}(x, w_s/T^s) \quad (9)$$

and the variance is given by

$$\sigma^2(x) = \frac{\sum_{s=1}^S \sum_{i \in A} [y_i - f_{NN}(x_i, w_s/T^s)]^2}{S - 1} \quad (10)$$

[15] The confidence interval (CI) at the  $\alpha\%$  significance level indicates that in repeated application of the technique, the frequency with which the CI would contain the true value is  $100 \times (1 - \alpha)\%$ . A typical value of  $\alpha$  is 0.05 which corresponds to  $(1 - 0.05) \times 100\% = 95\%$  confidence limits. A  $100 \times (1 - \alpha)\%$  CI covering the mean/ensemble demand  $\hat{y}(x)$  can be estimated using the following equation [Efron and Tibshirani, 1993]

$$CI = [UB, LB] = [\hat{y}(x) + t_{n-p}^{\alpha/2} \sigma(x), \hat{y}(x) - t_{n-p}^{\alpha/2} \sigma(x)] \quad (11)$$

where UB is the upper band, LB is the lower band,  $\sigma(x)$  is the standard deviation of  $S$  bootstrapped estimates,  $t_{n-p}^{\alpha/2}$  is the  $\alpha/2$  percentile for the Student  $t$  distribution with  $n - p$  degrees of freedom,  $n$  is the total number of water demand observations, and  $p$  is the total number of parameters in the NN model. A typical value of  $\alpha$  is 0.05.

## 3. Study Area and Data

[16] Montreal is the second largest city in Canada, with a population of approximately 1.6 million [Statistics Canada, 2007]. The city's six drinking water treatment plants draw water from three sources surrounding the city: Rivière des Prairies, Lake Sainte Louis, and the St. Lawrence River. The two treatment plants, Atwater and Charles-J. des Bailleurs, are some of the largest plants in Canada. They draw water from the St. Lawrence and have the capacity to treat 1364 million and 1136 million cubic meters of water per day, respectively, which comprises 88% of the entire water treatment capacity of Montreal, [City of Montreal, 2010]. Montreal's 5000 km water distribution network is old, with approximately 500 pipe breaks per year, some of which remain undetected for months or years. Studies in the past decade have indicated that approximately 40% of the treated water is lost, indicating a need to repair water lines and to improve water leak detection techniques. Since 2002, the city has begun major rehabilitation and repairs of the distribution system; it has been estimated that this will take a minimum of 20 years and cost several billions of dollars [City of Montreal, 2010]. On average, Montreal produces 2 million cubic meters of drinking water per day [City of Montreal, 2010]. Residents of the province of Quebec and Montreal consume more water than other

Canadians; one explanation for this has been the lack of metering within the province.

[17] In terms of weather, the average summer high in Montreal is 23°C, with a historic extreme high of 38°C, and the average winter low is -11°C, with a historic extreme low of -38°C. Annual rainfall in Montreal is almost 1000 mm, with an extreme daily rainfall of 93.5 mm. Annual snowfall averages over 200 cm with an extreme daily snowfall of 43.2 cm [Environment Canada, 2010]. The data that was used in the study was provided by the City of Montreal and consisted of average daily and average monthly water demand (WatDemand), maximum temperature (MaxT), and total precipitation (TotP).

**4. Model Development**

**4.1. Neural Network Structure Identification**

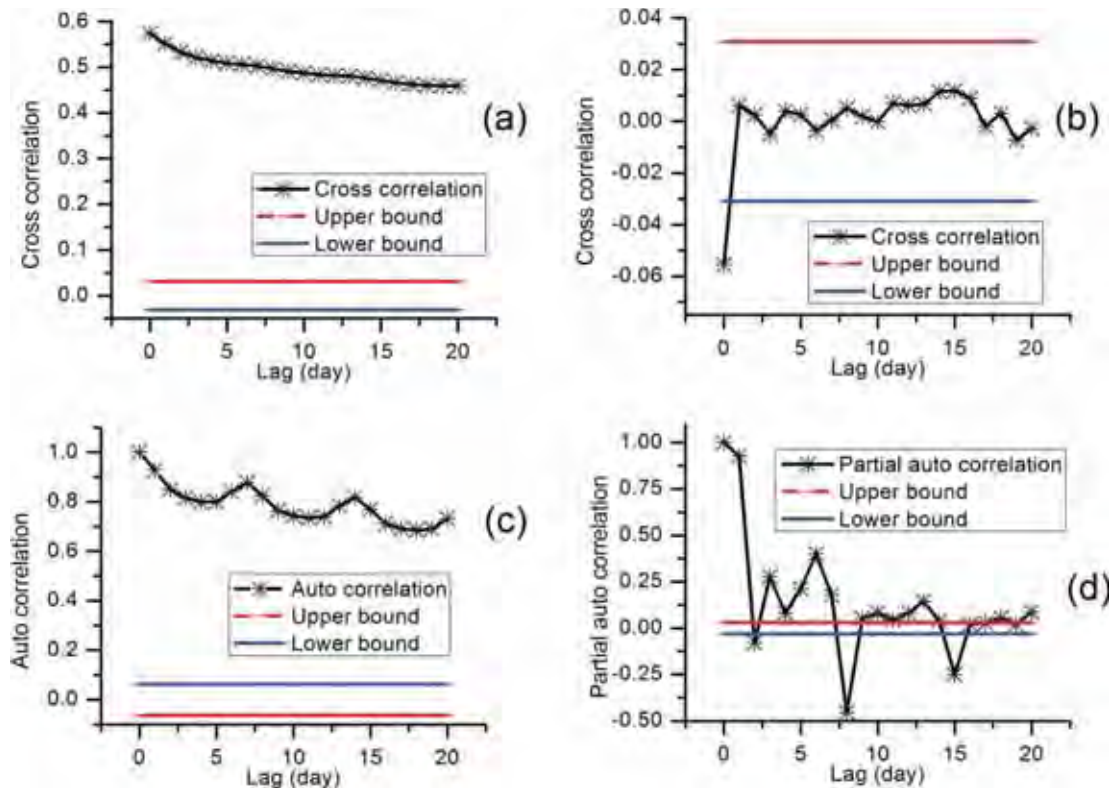
[18] Proper selection of input variables is critical in NN model development, as there is no direct method to select the optimum number of inputs. In this study, a trial and error method was used to select the optimum number of inputs along with a statistical approach recommended by Sudheer et al. [2002]. This statistical approach assumes that the important variables for each time lag can be identified by statistically analyzing the data series, by examining cross correlations, autocorrelations, and partial autocorrelations between the variables. Initially, this process was applied to

select significant inputs from the water demand data of Montreal for daily and monthly urban water demand forecasting. Significant inputs for weekly water demand forecasting were selected using the same cross-correlation statistics used for daily water demand time series. The cross-correlation statistics for the city of Montreal for daily time steps and monthly time steps are shown in Figures 1 and 2, respectively. Figures 1b and 2b show that precipitation does not play a significant role in determining water demand as the correlation between those two variables in Montreal is close to 0. Further, it is difficult to select significant inputs for maximum temperature and water demand lags using the cross-correlation statistics, because the correlation is not found to be significant. Thus, a trial and error procedure was adopted to determine the significant input variables. In this study, the number of hidden neurons that produced the lowest generalization error was determined to be the optimal structure [Jia and Culver, 2006]. For water demand forecasting in Montreal at different lead times, past information of water demand, total precipitation, and maximum temperature were considered, and the NN structures were tested for 1–14 hidden neurons.

**4.2. NN, BNN, WNN, WBNN, ARIMA, and ARIMAX Model Development**

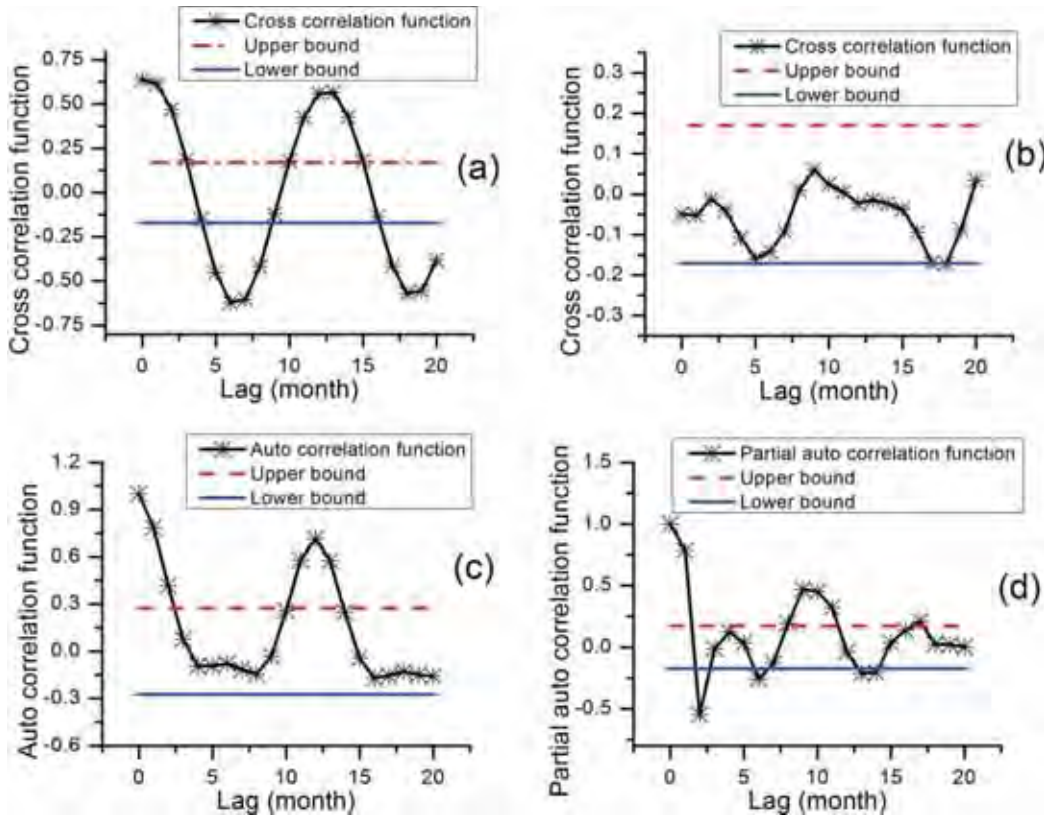
**4.2.1. NN Model Development**

[19] All the methodological issues need to be considered for the development of a robust NN model [Maier et al.,



**Figure 1.** Correlation functions of the daily time series data of Montreal: (a) cross-correlation function between daily water consumption and daily maximum temperature, (b) cross-correlation function between daily water consumption and daily total precipitation, (c) autocorrelation function of daily water consumption, and (d) partial autocorrelation function of daily water consumption.





**Figure 2.** Correlation functions of the average monthly time series data of Montreal: (a) cross-correlation function between monthly average water demand and monthly average maximum temperature, (b) cross-correlation function between monthly average water demand and average monthly total precipitation, (c) autocorrelation function of average monthly water demand, and (d) partial autocorrelation function of average monthly water demand.

2010]. Utmost care was taken to develop a robust NN model by selection of significant inputs such as variables and their lags, training with tracking on overfitting, selecting an appropriate training algorithm, and most importantly training the NN model while considering the generalization error [Jia and Culver, 2006]. In addition to regular NN models, three other NN-based models were developed in this study: bootstrap-based NN (BNN), wavelet-based NN (WNN), and wavelet-bootstrap-NN conjunction (WBNN) models. A multilayer perceptron (MLP) feed-forward NN model was initially developed using the most significant inputs, that were first log transformed and then linearly scaled to the range (0, 1) [Campolo *et al.*, 1999]. As discussed earlier, the data used in this study consisted of average daily and average monthly water demand (WatDemand), maximum temperature

(MaxT), and total precipitation (TotP) from 27 February 1999 to 6 August 2010. For the development of the different types of models, the data were divided into three sets: one for training the models, one for cross validation, and one for testing the performance of the developed models. The details of the data partitioning are shown in Table 1. The training data set was applied to train the ANN models by computing the gradient and updating the neural network connection weights and the biases. The testing data set applied to evaluate the performances of the models, whereas the cross-validation data set was applied for an early stopping approach to avoid overtraining or overfitting of the neural network models. In the early stopping approach, the objective function, the mean square error (MSE), is monitored at each iteration of training and cross validation simultaneously,

**Table 1.** Partitioning of Data for NN Model Development for Montreal

Data Set	Period	Number of Data Patterns for Daily and Weekly Water Demand Forecasting	Number of Data Patterns for Monthly Water Demand Forecasting
Training	2/27/1999 to 12/31/2007	3230	107
Cross-validation	1/1/2008 to 12/31/2008	366	13
Testing	1/1/2009 to 8/6/2010	583	18

and the training is stopped at the point the MSE for the cross-validation data reaches the minimum level. After this level, the NN model will start overfitting despite better performance (e.g., decreasing MSE) during training [Bishop, 1995]. To increase computational efficiency, a second-order training method, the Levenberg-Marquardt method, was used to minimize the mean squared error between the forecasted and observed water demands. All four NN-based models were developed with Matlab codes using MATLAB® (v.7.10.0), except for the generation of realizations of the training data set to develop BNN and WBNN models, which was done using an Excel add-in (Bootstrap.xla) [Barreto and Howland, 2006].

[20] The NN models (i.e., NN, BNN, WNN, and WBNN) were developed using daily average water demand, daily maximum temperature, daily average total precipitation for 1, 3, and 5 day lead time water demand forecasting. For weekly water demand forecasting, instead of weekly average water demand, weekly maximum temperature, and weekly average total precipitation for 1 and 2 week lead times, daily average water demand, and daily maximum temperature for 1 and 2 week (i.e., 7 and 14 day) lead times was forecasted. It should be noted that when we refer to weekly forecasting in this study, the aim was to forecast for exactly the same day the following week. This was done since it was deemed to be more useful to forecast water demand for the same day 1 week ahead (and the same day 2 weeks ahead) given that water demand varies depending on the day (for example, whether it is a week day or a weekend day), instead of forecasting total average 1 week and 2 week water demand. Monthly water demand for 1 and 2 months was forecasted using monthly average water demand, monthly average maximum temperature, and monthly average total precipitation.

#### 4.2.2. WNN Model Development

[21] Next, the WNN model was developed by inputting the wavelet subtime series, produced using the discrete wavelet components (DWCs) at varying scales, into the NN model. Each of these subtime series plays a distinct role in describing the original water demand series. The wavelet functions used in this research are Haar, Daubechies (i.e., db2, db3, db4, db5, db6), Sym3, and Coif1 [Nourani et al., 2009; Wu et al., 2009], and 1–5 levels of decomposition were considered in this study. As the performance of the db5 function derived from the family of Daubechies wavelets with three levels of decomposition was found to be the best, for illustration purposes three levels of decomposition (d1, d2, and d3) and approximation (A3) for the daily and monthly water demand and temperature data of Montreal are shown in Figures 3 and 4, respectively. As explained earlier, it should be noted that decomposed daily time series data were also used to develop models for weekly water demand forecasting with 7 and 14 day (i.e., 1 week and 2 week) lead times, respectively. The effective wavelet subtime series were determined using the correlation coefficients between each wavelet component and the observed water demand. Table 2 shows the correlation between the original daily and monthly time series and the corresponding different wavelet subtime series for Montreal. In earlier studies [Tiwari and Chatterjee, 2010b, 2011; Kisi, 2010; Adamowski and Sun, 2010], the significant wavelet subtime series of a particular time series were added and used, which became the

new inputs to develop the WNN model. In this study, in contrast to previous studies, considering that all the wavelet subtime series may play a significant role in the original time series, instead of selecting on the basis of a particular threshold level, all the components were given due consideration to evaluate their effectiveness to forecast water demand in the city of Montreal. The performance of the developed models was evaluated using five performance indices, namely: coefficient of determination ( $R^2$ ), root-mean-square error (RMSE), percentage deviation in peak ( $P_{dv}$ ), mean average error (MAE), and persistence index (PI). Persistence index (PI) is one minus the ratio of the sum square error to what the sum square error would have been if the forecast had been the last observed value.

#### 4.2.3. BNN Model Development

[22] The BNN model was developed as an ensemble of 100 NNs built using bootstrap resamples of raw data sets (i.e., the significant input variables identified when developing the NN models), whereas the WBNN model was developed as an ensemble of 100 NNs built using bootstrap resamples of wavelet subtime series (i.e., the significant input variables identified when developing the WNN models), instead of raw data sets. Thus, for a given lead time, there were 100 forecasts from a single testing data set or, in other words, by using the bootstrap technique, each lead time had 100 sets of weights instead of one. The 100 forecasted values for each lead time were used to build 95% confidence bands that depict the uncertainty associated with the forecasts.

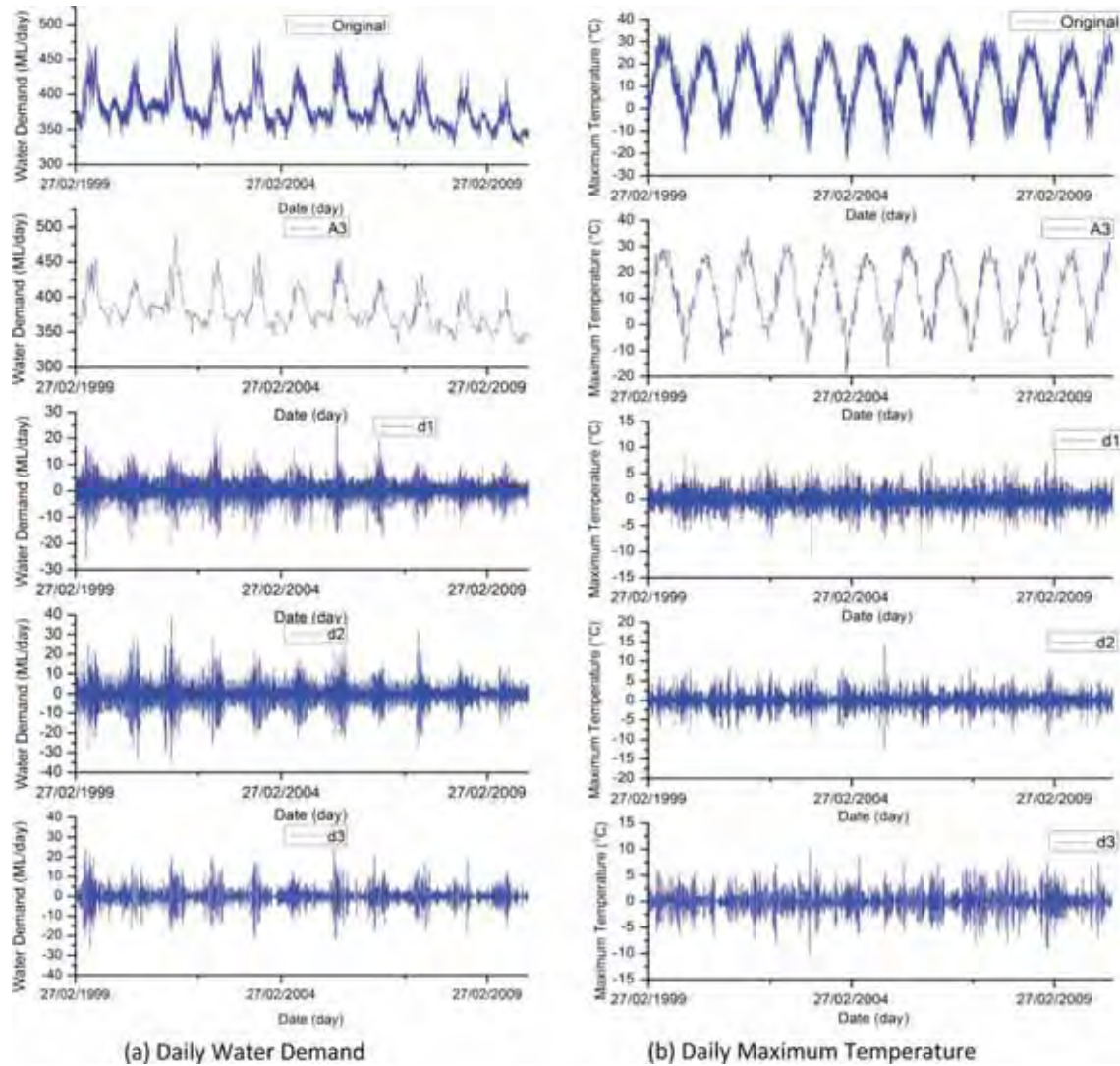
#### 4.2.4. WBNN Model Development

[23] The WBNN model takes advantage of the capabilities of both the bootstrap resampling and wavelet transform techniques. To maintain consistency with the BNN model, the WBNN model was also developed using 100 resamples of the same significant input variables identified when developing the WNN models. Bootstrap.xla, an Excel add-in [Barreto and Howland, 2006], was used to generate bootstrap resamples of raw data sets for the BNN models and wavelet subtime series for the WBNN models. The performance of the best model was also tested using different numbers of bootstrap resamples (i.e., 25, 50, 200, and 500) to evaluate the effectiveness of the number of bootstrap resamples used. The BNN and WBNN models were used to assess and quantify the uncertainty associated with the 1, 3, and 5 day, 1 and 2 week, and 1 and 2 month lead time forecasts by developing confidence bands using the ensemble forecasts.

#### 4.2.5. ARIMA and ARIMAX Model Development

[24] ARIMA and ARIMAX models (i.e., ARIMA models with additional independent input variables) were developed to forecast water demand in the city of Montreal as a benchmark to evaluate the performance of the different NN models and were developed using the SPSS software package (version 10, SPSS Inc., Chicago, Illinois). Initially, the stationarity of the input data series was determined by the autocorrelation function (ACF). It was observed that urban water demand data from Montreal was nonstationary. Therefore, the data sets for ARIMA and ARIMAX modeling were transformed into a stationary time series through the differencing process. The development of the ARIMA models in this study followed the methodology used by Adamowski [2008, 2012].





**Figure 3.** Wavelet subtime series of the (a) daily water demand and (b) daily maximum temperature of Montreal from 27 February 1999 to 6 August 2010.

## 5. Results and Discussion

### 5.1. Daily Water Demand Forecasting

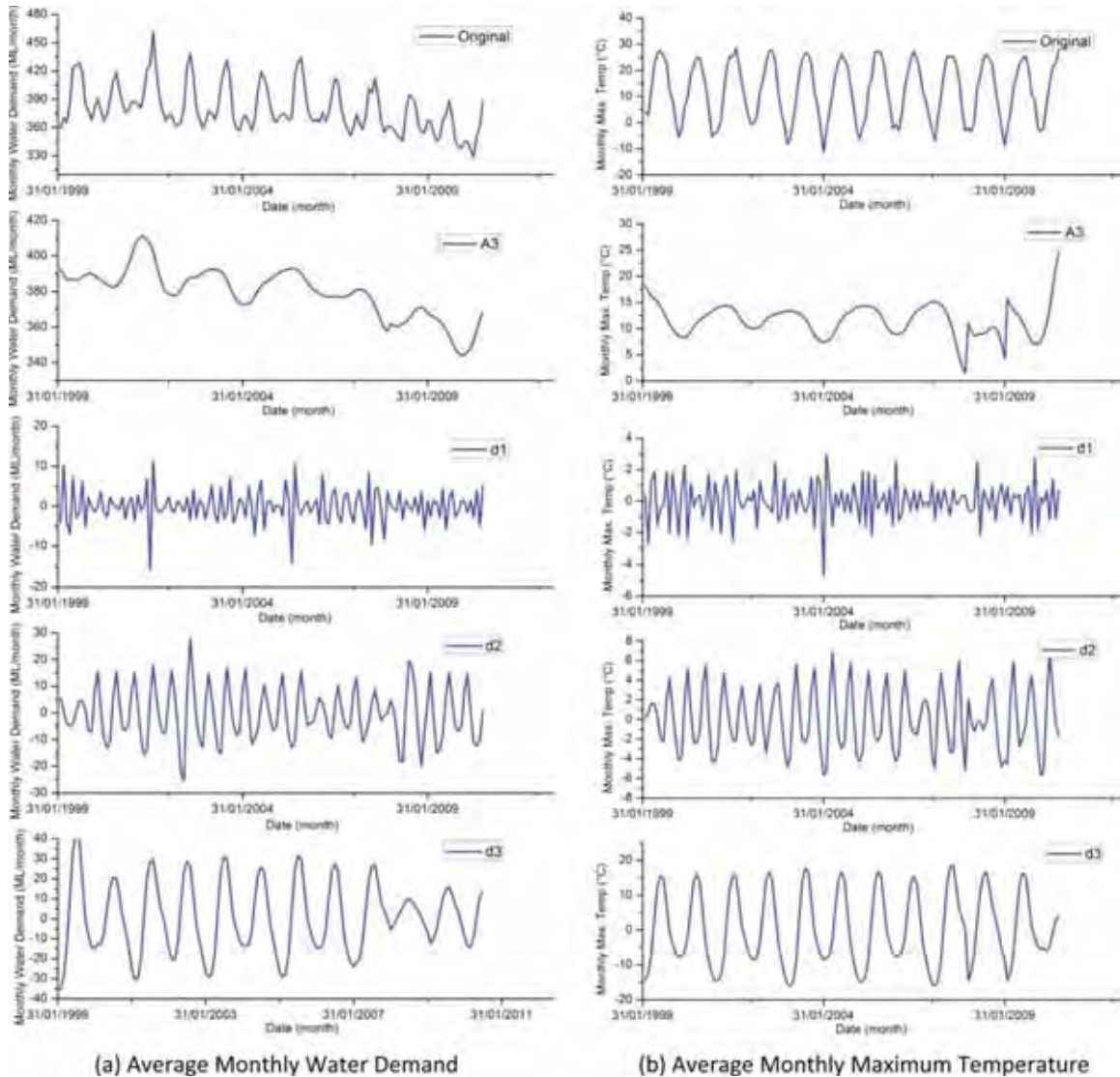
#### 5.1.1. Daily Water Demand Forecasting Using NN Models

[25] The performance of the best NN model is presented in terms of different performance indices for 1, 3, and 5 day lead times in Table 3 and Figure 5. Daily precipitation and daily maximum temperature were not found to have a significant impact on daily lead time water demand forecasting. The daily time step performance is satisfactory up to 5 day lead time forecasts in terms of the different performance indices. For 1 day lead time forecasts, it can be observed from the scatter plots that the performance is good for low, medium, and high demand profiles, but for 3 and 5 days the performance deteriorates significantly for higher demand profiles. It is obvious that for higher lead times, model performance deteriorates, as there is less in-

formation available for longer lead time horizons, but the deteriorating performance specifically for higher values at longer lead times shows the weakness of the NN model structure to forecast the higher water demand values in Montreal. Another reason may be that the number of water demand values is lower for high water demand values compared to low and medium water demand values, and the NN models try to dampen the high and medium demand values and underestimate the values. This indicates that NN models are not very effective in capturing the nonstationarity in the data and show the weakness of the NN model structure.

#### 5.1.2. Daily Water Demand Forecasting Using BNN Models

[26] To improve the performance of NN models, BNN models were developed by generating ensemble forecasts of water demand in Montreal for 1, 3, and 5 day lead times. Considering the range of variation of water demand in



**Figure 4.** Wavelet subtime series of the (a) monthly average water demand and (b) monthly average maximum temperature of Montreal from February 1999 to July 2010.

Montreal during the testing period from a high of 431.67 ML/d to a low of 318.12 ML/d, the BNN model with RMSE values of 6.06, 10.70, and 12.55 ML/d performed well for 1, 3, and 5 day lead times (Table 3 and Figure 5), respectively. The BNN model simulated the higher water demand values better than NN models for 3 and 5 day lead times forecasts. It can be observed from the scatter plots that compared to the best NN model, for 1, 3, and 5 day lead times the BNN model performed better for higher water demand values. However, the BNN model underestimated several higher water demand values, especially for higher lead time forecasts. The BNN model has the capability to produce more stable solutions by producing ensemble forecasts; however, it is not able to extract the nonstationarity from the data set. Considering the need to improve model performance, WNN models were developed.

### 5.1.3. Daily Water Demand Forecasting Using WNN Models

[27] For daily water demand forecasting, the performance of WNN models in terms of  $R^2$ , RMSE, PI, and MAE for all 1, 3, and 5 day lead time forecasts is much better compared to the best NN and BNN models (Table 3 and Figure 5). The better performance of the WNN model may be due to the reason that NN and BNN models have limitations to extract the nonstationarity and the physical structure from the training data set. It can be observed from the scatter plots that the WNN model simulates the observed values very well compared to the NN and BNN model forecasts. Wavelet analysis simplifies the physical structure of the data, simplifying the learning process of WNN models during training. This allows for better simulation of the observed values, even for higher water demand values whose numbers are much lower in the training data set. It

**Table 2.** Correlations Between Different Wavelet Subtime Series and the Original Water Demand Time Series in Montreal

Wavelet Subtime Series	Daily		
	WatDmand	MaxT	TotP
A3	0.94	0.58	-0.08
D1	0.16	0.03	-0.02
D2	0.25	0.06	-0.01
D3	0.19	0.07	-0.02
Original	1.00	0.58	-0.06
		Monthly	
A3	0.55	0.20	-0.09
D1	0.18	0.04	0.00
D2	0.42	-0.11	-0.08
D3	0.69	0.65	0.10
Original	1.00	0.64	-0.05

can be noted that the BNN model produces more stable forecasts by combining forecasts made using different realizations of the training data set, whereas the WNN model reduces noise and extracts nonstationarity from the training data set. To further enhance the performance of the NN model, a hybrid WBNN model was developed by combining the strength of BNN and WNN models.

**Table 3.** Performance of the Best Models Using NN, BNN, WNN, WBNN, ARIMA, and ARIMAX Models for 1, 3, and 5 Day Lead Time Water Demand Forecasting in Montreal<sup>a</sup>

Lead Time	Best Model Structure		Performance Indices					
			Hidden Neurons (HN)	R <sup>2</sup>	RMSE (ML/d)	P <sub>dv</sub> (%)	MAE (ML/d)	PI
			<i>NN</i>					
1 day	WatDmand(t), WatDmand(t-1)	WatDmand(t-2), WatDmand(t-3) WatDmand(t-4) WatDmand(t-5) WatDmand(t-6), WatDmand(t-7), WatDmand(t-8) WatDmand(t-13)	13	0.92	5.88	0.88	4.17	0.45
3 day	WatDmand(t), WatDmand(t-1)	WatDmand(t-2), WatDmand(t-3) WatDmand(t-4) WatDmand(t-5) WatDmand(t-6), WatDmand(t-7)	5	0.77	9.92	2.52	7.15	0.41
5 day	WatDmand(t), WatDmand(t-1)	WatDmand(t-2), WatDmand(t-3) WatDmand(t-4) WatDmand(t-5) WatDmand(t-6), WatDmand(t-7), WatDmand(t-8)	11	0.71	10.91	4.80	7.11	0.37
			<i>BNN</i>					
1 day	Same inputs as in NN			0.90	6.09	1.66	4.41	0.41
3 day	Same inputs as in NN			0.76	10.70	2.42	7.96	0.31
5 day	Same inputs as in NN			0.64	12.55	0.07	9.14	0.07
			<i>WNN</i>					
1 day	a3(t), d3(t), d2(t), d1(t) components of WatDmand(t)			0.98	3.11	-0.57	2.21	0.84
3 day	a3(t), d3(t), d2(t), d1(t) components of WatDmand(t)			0.94	5.33	-0.82	4.17	0.83
5 day	a3(t), d3(t), d2(t), d1(t) components of WatDmand(t)			0.90	7.81	0.37	6.30	0.68
			<i>WBNN</i>					
1 day	Same inputs as in WNN			0.98	4.09	-0.97	2.99	0.73
3 day	Same inputs as in WNN			0.96	6.08	1.04	4.81	0.78
5 day	Same inputs as in WNN			0.90	7.29	0.54	5.82	0.72
			<i>ARIMA</i>					
1 day				0.90	6.51	0.58	4.85	0.33
3 day				0.70	10.55	1.32	7.60	0.30
5 day				0.59	14.57	0.79	11.33	0.12
			<i>ARIMAX</i>					
1 day				0.91	6.55	2.70	4.72	0.39
3 day				0.71	10.42	1.4	7.44	0.37
5 day				0.64	11.44	2.25	8.83	0.41

<sup>a</sup>WatDmand (t + 1) = 1 day/week/month lead water demand forecast; WatDmand(t) = Total water demand at time (t).

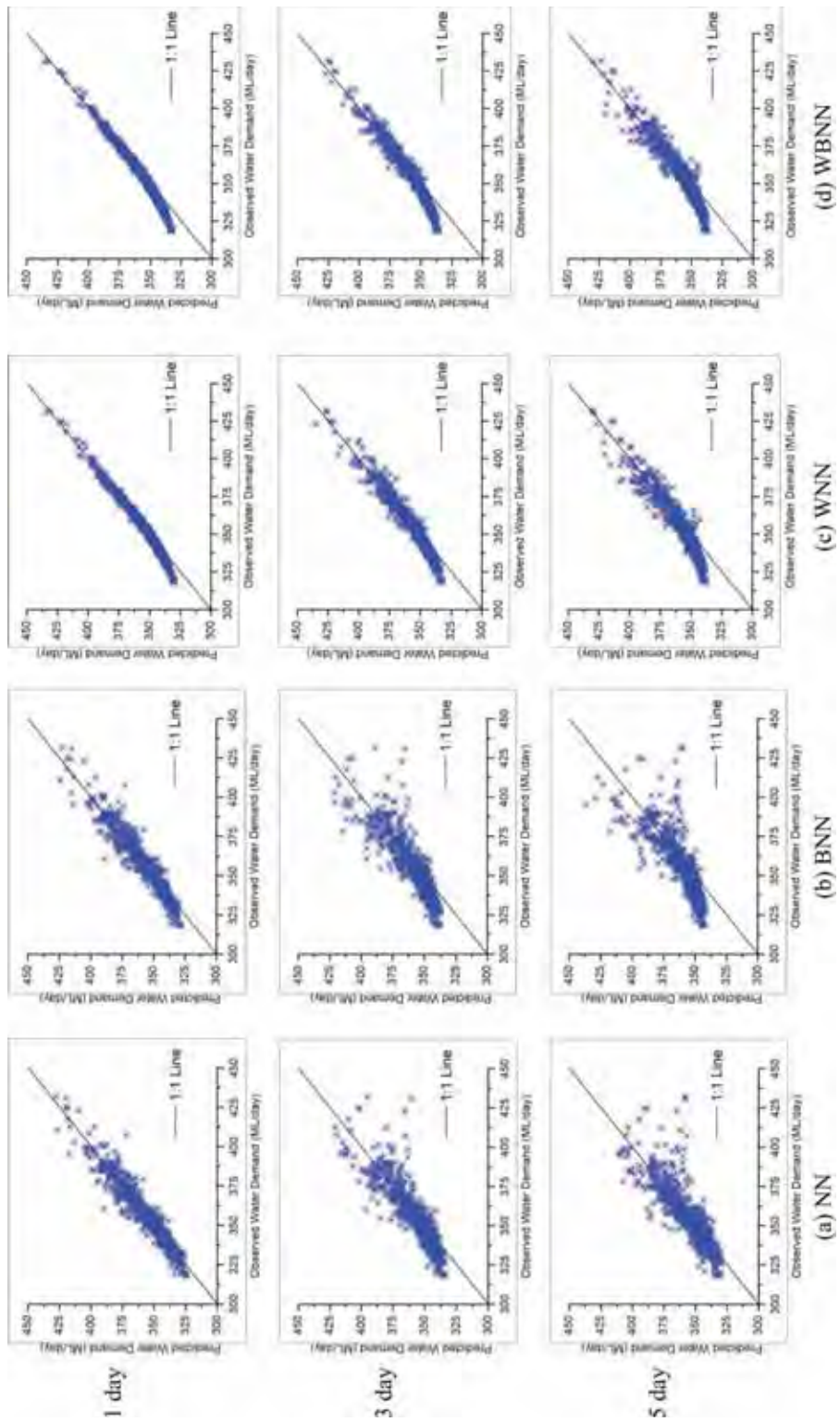
**5.1.4. Daily Water Demand Forecasting Using WBNN Models**

[28] The WBNN models, which use the capabilities of both the WNN and BNN models, performed very well (Table 3 and Figure 5). The performance of the best WNN and WBNN models was found to be better compared to the best NN and BNN models. The performance of the best WNN model was found to be better than the WBNN model for 1 and 3 day lead times, whereas the WBNN model performed better for 5 day lead time forecasts. The overall performance of the WBNN models was considered better compared to the best WNN models. The forecasts obtained using the WBNN models are more accurate as the bootstrap technique reduces the variance, and wavelet analysis reduces the noise, making the periodic information more obvious. Further, the different forecasts obtained from the WBNN model can be used to develop confidence bands to assess the uncertainty associated with the forecasts.

**5.1.5. Daily Water Demand Forecasting Using ARIMA and ARIMAX Models**

[29] It was observed that the inclusion of additional inputs, namely temperature and precipitation, by developing ARIMAX models does not improve the performance in terms of different performance indices (Table 3), since the





**Figure 5.** Scatter plots for observed and predicted water demand in Montreal for 1 day, 3 day, and 5 day lead time forecasts for the testing data set using: (a) NN, (b) BNN, (c) WNN, and (d) WBNN models.

performance of both ARIMA and ARIMAX models are very similar, with the ARIMAX model performing slightly better than the ARIMA model for 3 and 5 the day lead time forecasts. These exogenous inputs may increase the model performance for longer lead times. Overall, for daily forecasting (1, 3, and 5 days), the performance of the WBNN models was found to be more accurate and reliable than the NN, BNN, WNN, ARIMA, and ARIMAX models.

**5.2. Weekly Water Demand Forecasting**

**5.2.1 Weekly Water Demand Forecasting Using NN Models**

[30] Weekly water demand forecasting for 1 week and 2 week lead times was carried out using NN, BNN, WNN, WBNN, ARIMA, and ARIMAX models, and the performance of the different models is presented in Table 4 and Figure 6. It can be observed in the case of weekly water demand forecasts that daily precipitation and daily maximum temperature play a significant role and improved model performance. Performance of the traditional NN models in terms of different performance indices for 1 week lead time water demand forecasting can be considered satisfactory. Further, even though the observed and forecasted values using the NN model are very close for 2 week lead time forecasts, lower and medium values are overestimated whereas peak values are underestimated. This shows that the performance of the NN model is not acceptable for 2 week lead time forecasts. The performance in terms of persistence index (PI) shows that even though the NN model can forecast better than a simple naïve persistence model, there is a need to apply hybrid approaches

to improve the performance of NN models for higher step weekly water demand forecasting.

**5.2.2 Weekly Water Demand Forecasting Using BNN Models**

[31] The performance of BNN models was found to be very close to NN models, and both the models lacked generalization capabilities (Table 4 and Figure 6) as several values were overestimated or underestimated (especially for 2 week lead time forecasts). This reflects the inability of NN models to capture the nonstationarity from the input and output variables for longer lead time forecasts. The significant deviation from the 1:1 line for different water demand values shows that NN and BNN models have limited capability to extract nonstationarity from the data set, and lack generalization ability as peak values are dampened toward frequently occurring lower and medium water demand values.

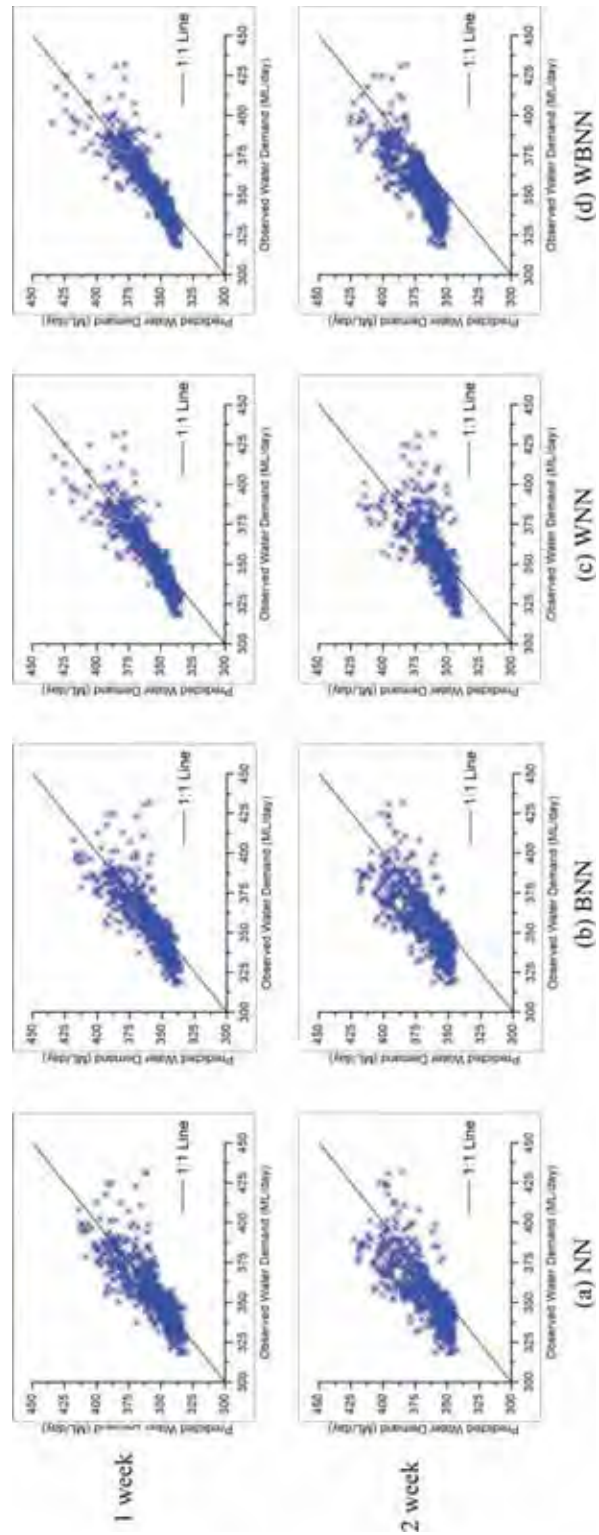
**5.2.3 Weekly Water Demand Forecasting Using WNN Models**

[32] The performance of WNN models was tested for 1 week and 2 week lead time forecasts and the performance in terms of different performance indices and scatter plots (Table 4 and Figure 6) showed that WNN model forecasts for both 1 and 2 week lead times is much better compared to the best NN and BNN models. The best WNN model performed well, since the forecasted values are very close to the 1:1 line, even for the 2 week lead time water demand forecasts. Performance of the WNN model was significantly better than the NN model for 2 week lead time forecasts; this shows that for longer lead time forecasts that are highly affected by nonstationarity (i.e., trends and

**Table 4.** Water Demand Forecasting in Montreal for 1 and 2 Week Lead Times using NN, BNN, WNN, WBNN, ARIMA, and ARIMAX Models

Lead Time	Best Model Structure	HN	R <sup>2</sup>	RMSE (ML/d)	P <sub>dv</sub> (%)	MAE (ML/d)	PI
<i>NN</i>							
1 week	WatDmand(t), WatDmand(t-6), MaxT(t), MaxT(t-6), TotP(t), TotP(t-6)	5	0.67	11.83	4.07	8.59	0.12
2 week	WatDmand(t), WatDmand(t-6), MaxT(t), MaxT(t-6), TotP(t), TotP(t-6)	8	0.59	17.59	2.09	14.45	0.17
<i>BNN</i>							
1 week	Same inputs as in NN		0.68	12.33	3.15	9.17	0.10
2 week	Same inputs as in NN		0.57	18.02	2.75	15.22	0.12
<i>WNN</i>							
1 week	a3(t), d3(t), d2(t), d1(t) of TotCons; A3 and d3 of MaxT and A3 of TotP with 1,2 and 3 day lag time variables	5	0.76	12.07	-1.76	8.90	0.35
2 week	a3(t), d3(t), d2(t), d1(t) of TotCons; A3 and d3 of MaxT and A3 of TotP with 1 and 7 day lag time variables	3	0.72	15.49	-0.14	10.93	0.28
<i>WBNN</i>							
1 week	Same inputs as in WNN		0.78	10.06	-0.58	7.49	0.46
2 week	Same inputs as in WNN		0.68	18.76	1.47	16.28	0.23
<i>ARIMA</i>							
1 week			0.68	11.74	0.03	7.49	0.10
2 week			0.52	16.95	4.36	14.46	0.15
<i>ARIMAX</i>							
1 week			0.68	11.89	2.52	8.80	0.09
2 week			0.51	17.03	3.90	14.58	0.14





**Figure 6.** Scatter plots for observed and predicted water demand in Montreal for 1 week and 2 week lead time forecasts for the testing data set using: (a) NN, (b) BNN, (c) WNN, and (d) WBNN models.

seasonality), wavelet analysis can be used to improve the model performance.

**5.2.4. Weekly Water Demand Forecasting Using WBNN Models**

[33] The performance of the best WNN and WBNN models was better compared to the best NN and BNN models (Table 4 and Figure 6). The performance of the best WNN model was found to be better than the WBNN model for 2 week lead times, whereas the WBNN model performed better for 1 week lead time forecasts. The overall performance of the WBNN models was considered better compared to the best WNN models. The forecasts obtained using the WBNN model are more accurate than NN, BNN, WNN, ARIMA, and ARIMA models because the bootstrap technique reduces the variance, and wavelet analysis reduces the noise and makes the periodic information more easily understandable by the model. The WBNN model provides more reliable forecasts since they are obtained using different realizations of the training data set that averages over the error. Further, similar to BNN models, the different forecasts obtained from the WBNN models can be used to develop confidence bands to assess the uncertainty associated with the forecasts.

**5.2.5. Weekly Water Demand Forecasting Using ARIMA and ARIMAX Models**

[34] Compared to daily water demand forecasting where inclusion of exogenous variables (i.e., daily maximum temperature and daily total precipitation) in the ARIMAX model improved the model performance compared to the ARIMA model (Table 4), the inclusion of maximum temperature and total precipitation in the ARIMAX model does not improve the model performance for 1 and 2 week lead time water demand forecasting. Performance of the

ARIMA and ARIMAX models were very close to that of the NN and BNN models; however, their performance was significantly worse than the WNN and WBNN models.

**5.3. Monthly Water Demand Forecasting in Montreal**

**5.3.1. Monthly Water Demand Forecasting Using NN Models**

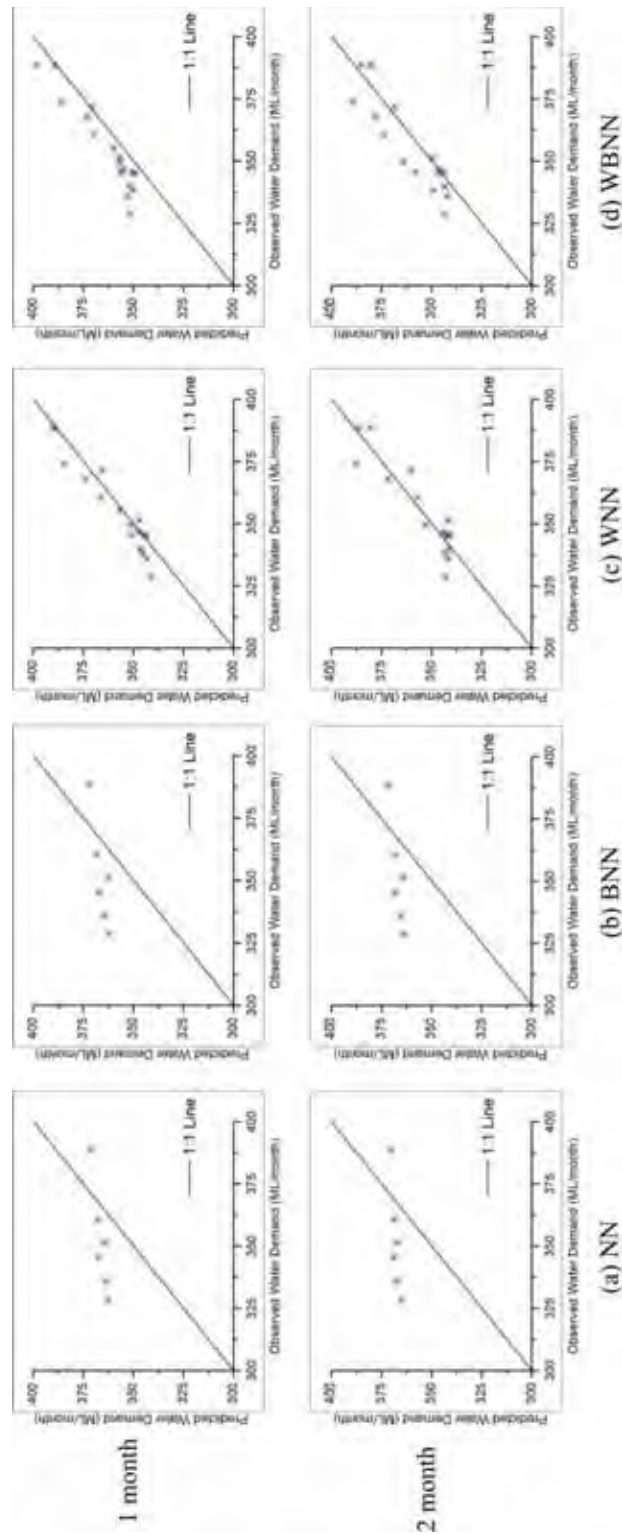
[35] Monthly water demand forecasting for 1 and 2 month lead times was carried out using NN, BNN, WNN, WBNN, ARIMA, and ARIMAX models, and it was observed that in monthly average water demand forecasting, monthly average total precipitation, and monthly average maximum temperature play a significant role and improved model performance (Table 5 and Figure 7). Considering the range of variation of average monthly water demand in Montreal (a high of 388.6 ML/month to a low of 328.7 ML/month), the performance of the regular NN model for monthly time steps and with a RMSE value of 22.18 and 24.27 ML/month cannot be considered satisfactory for 1 and 2 month lead times, respectively. It can also be observed from the scatter plots that regular NN models are not able to simulate the monthly average water demand values satisfactorily since the values diverge from the 1:1 line. Moreover, the lower values are overestimated and the higher value is underestimated.

**5.3.2. Monthly Water Demand Forecasting Using BNN Models**

[36] The results for forecasts of water demand in Montreal for 1 and 2 month lead times show that the performance of the BNN models is very similar to the NN models (Table 5 and Figure 7), and the performance of both the NN and BNN models cannot be considered satisfactory in terms of the different performance indices. It can also be seen from the scatter plots that the BNN model does not

**Table 5.** Monthly Water Demand Forecasting in Montreal for 1 and 2 Months Lead Times Using NN, BNN, WNN, WBNN, ARIMA, and ARIMAX Models

Lead Time	Best Model Structure	HN	R <sup>2</sup>	RMSE (ML/month)	P <sub>av</sub> (%)	MAE (ML/month)	PI	
		<i>NN</i>						
1 month	WatDmand(t), WatDmand(t-11), Maxt(t), Maxt(t-11), TotP(t), TotP(t-11)	2	0.87	22.18	4.51	20.31	0.06	
2 month	WatDmand(t), WatDmand(t-10), Maxt(t), Maxt(t-11), TotP(t), TotP(t-11)	8	0.70	24.27	4.69	22.33	0.54	
		<i>BNN</i>						
1 month	Same inputs as in NN		0.77	21.74	3.84	19.70	0.10	
2 month	Same inputs as in NN		0.74	23.03	4.34	20.97	0.58	
		<i>WNN</i>						
1 month	A3, d1, d2, d3 components of WatDmand(t), A3 and d3 components of MaxT(t) and d3 component of TotP(t)	3	0.96	5.94	-0.33	4.82	0.81	
2 month	A3, d1, d2, d3 components of WatDmand(t), A3 and d3 components of MaxT(t) and d3 component of TotP(t)	5	0.91	7.21	0.23	5.99	0.89	
		<i>WBNN</i>						
1 month	Same inputs as in WNN		0.77	11.13	-0.68	8.62	0.32	
2 month	Same inputs as in WNN		0.54	13.49	2.25	10.94	0.61	
		<i>ARIMA</i>						
1 month			0.57	16.98	4.29	14.86	0.47	
2 month			0.67	19.68	2.99	17.70	0.69	
		<i>ARIMAX</i>						
1 month			0.90	19.20	-4.06	18.21	0.32	
2 month			0.90	18.62	-3.94	17.61	0.72	



**Figure 7.** Scatter plots for observed and predicted daily water demand in Montreal for 1 month and 2 month lead time forecasts for the testing data-sets using : (a) NN, (b) BNN, (c) WNN, and (d) WBNN models.

simulate the water demand values satisfactorily as the simulated values diverge from the 1:1 line. This again illustrates that NN models lack the ability to extract nonstationarity from the training data set.

### 5.3.3. Monthly Water Demand Forecasting Using WNN Models

[37] WNN models were also used to forecast water demand in Montreal for 1 and 2 month lead times, and the performance was much better compared to the best NN and BNN models (Table 5 and Figure 7). The performance of the WNN models with a monthly time step and with a RMSE value of 5.94 and 7.21 ML/month can be considered very good for 1 and 2 month lead times, respectively. It can also be observed that the WNN model simulated the values very well as the values are very close to the 1:1 line. More importantly, all the values (high and low) are simulated very well and there is no significant evidence of underestimation and overestimation of water demand values as can be observed from scatter plots.

### 5.3.4. Monthly Water Demand Forecasting Using WBNN Models

[38] In the case of monthly water demand forecasting, the WBNN model with a RMSE value of 13.49 ML/month performed well up to a 2 month lead time forecast (Table 5 and Figure 7). The performance of the best WNN and WBNN models was better compared to the best NN and BNN models, with the best WNN model providing better results than the WBNN model. However, overall the performance of the WBNN models can be considered better compared to the best WNN models. The forecasts obtained using the WBNN model are more accurate because the bootstrap technique reduces the variance, and wavelet analysis reduces the noise and makes the periodic information more readily understandable by the model.

### 5.3.5. Monthly Water Demand Forecasting Using ARIMA and ARIMA Models

[39] The performance of the ARIMA and ARIMAX models was found to be better than the NN model and BNN models, but not better than the WNN and WBNN models (Table 5). The better performance of the ARIMAX model compared to the ARIMA model for a 2 month lead time shows the significance of exogenous inputs for modeling longer lead time water demand forecasts.

## 5.4. Discussion

### 5.4.1. Comparative Performance of the Models

[40] It was observed that NN models are not able to extract the nonstationarity from the data set. Moreover, the NN models demonstrated their weakness to extract the nonlinearity when the length of the training data set is small, as it was noticed that the length of the data set is too small (i.e., net 107 data patterns) in NN, BNN, WNN, and WBNN models for 1 and 2 month lead time forecasts for training compared to the data patterns used for daily and weekly water demand forecasting (i.e., net 3230 data patterns in NN, BNN, WNN and WBNN models for 1,3, and 5 day; and 1 and 2 week lead time forecasts for training). Due to the higher number of data patterns the NN and BNN models simulated the observed values much better for the daily and weekly lead time forecasts; the performance of the NN and BNN models was not very good for monthly lead time forecasts. This may be due to the dominance of

nonstationarity in monthly water demand time series. This shows that the NN models and BNN models (which are ensembles of different NN models) have limited capability to extract the nonstationarity from the data sets. Further, it can be observed that the performance of WNN and WBNN models (which are the ensembles of several WNN models) are much better compared to NN and BNN models. This is because wavelet transform decomposed components of time series data extract different time varying components (i.e., trends and nonstationarity) (Figures 3 and 4) that may be representative of the sum of the subprocesses associated with the original time series data set. These different components facilitate the ability of the NN models that use WTs (i.e., WNN and WBNN) to extract nonlinearity and nonstationarity, and therefore their performance is superior to NN models developed using raw data sets.

[41] The results show the advantages of using wavelets in NN models, especially as the lead time increases. The models developed without wavelet transformed data consistently underestimated higher water demand values, especially at longer lead times. Accurately forecasting these high peak demands is operationally very important. Although the use of bootstrapping did not improve model performance, overall its additional capacity to reduce uncertainty is considered advantageous, outweighing the slight reduction in performance seen at some lead times. The performance of the WBNN method was comparable to the WNN method for 1 and 3 day lead time water demand forecasts, and the WBNN model performed slightly better for 5 day lead time forecasts. For 1 and 2 week water demand forecasting performance, the WBNN and WNN models, respectively, were found to be the best compared to the remaining models. For of 1 and 2 month water demand forecasting, the performance of the WNN model was found to be the best compared to the remaining models.

[42] It should be noted that in some of the cases the WNN models performed better than WBNN models, but the WBNN models are more consistent and reliable compared to WNN models, considering that they are an ensemble of several WNN models developed using different realizations of the training data set, and thus provide forecasts with reduced variance (i.e., reduced uncertainty). The narrow confidence bands of the WBNN models, along with the higher number of values inside the confidence bands enables the WBNN models (for daily, weekly, and monthly water demand forecasting) to be the most reliable method compared to the other methods. This is due to the reason that the WBNN models use the capability of wavelet analysis, which reduces the noise, and bootstrap resampling, which reduces the variance. The performance of the WBNN models is considered more reliable and accurate than the WNN models, even though in some cases the WNN models performed slightly better than the WBNN models. The reason is that WBNN models are an ensemble of WNN models developed using 100 realizations of the training data set, and not merely an ensemble of some selected better performing WNN models out of these 100 forecasts. WBNN models developed using different realizations of wavelet subtime series data may be representing different processes associated with water demand simulations at different time-frequency domains, and the WBNN models average over the error and produce more accurate and reliable forecasts.



[43] The performance of ARIMA and ARIMAX models were found to be better than NN and BNN models for weekly and monthly water demand forecasting. The better performance of ARIMA and ARIMAX models compared to NN and BNN models for weekly and monthly forecasts may be because NN models are not very effective at capturing nonstationarity in the data set, and longer time step (i.e., weekly and monthly) time series data shows trend and nonstationarity, as can be observed in Figure 2. ARIMA and ARIMAX models are able to extract this nonstationarity from the data set and provide better forecasts than NN and BNN models. However, after decomposing the data set using wavelet analysis by extracting the trend and nonstationarity, the performance of WNN and WBNN models improved significantly as discussed earlier. Moreover, the computation time using WBNN models proposed in this study is approximately 3–5 min using average system configurations such as an Intel Core i5 processor with 3GB RAM. This can be considered to be a time efficient process, and the proposed WBNN model can readily be implemented for operational water demand forecasting.

#### 5.4.2. Effect of Input Variables on Water Demand Forecasts

[44] In this study, it was observed that the performance of traditional NN models does not improve for daily water demand forecasting by including additional parameters (i.e., max temperature and total precipitation), whereas it improves for weekly and monthly lead time forecasts. It can be observed from the study that to forecast daily water demand values only previous water demand values are relevant, whereas for weekly and monthly water demand forecasting, average maximum temperature, and average total precipitation play a significant role. The reason may be that daily water demand values are more highly correlated with the previous water demand values than the previous maximum temperature and total precipitation values, and as such it is easier to simulate the short lead time water demand values using only the previous lagged values of water demand itself, whereas for weekly water demand values where autocorrelation for water demand values is weak, temperature and precipitation play a significant role in extracting relevant information. Similarly, in the case of monthly average water demand forecasting, maximum temperature and total precipitation play a significant role and the performance of the WNN and WBNN models improved. This may be due to the reason that water demand does not depend on the sudden occurrence of heavy rainfall or sudden variation in temperature (i.e., why daily time step maximum temperature and total precipitation have no effect on daily water demand forecasts), but it depends on the general/average weather conditions occurring in the region. It can also be supported by the observation that for 1 and 2 month average water demand forecasting, only the previous 1 month average condition of maximum temperature, average precipitation, and average water demand values are significant (Table 5), whereas in the case of daily and weekly water demand forecasting previous values for several lag time steps are required. This indicates that the daily and weekly water demand models are dependent on average conditions of lag time and not only discrete or single values of previous day water demand, maximum temperature, and total precipitation. This again supports the

idea that these water demand forecasts are dependent on average weather conditions and not on the sudden occurrence of heavy rainfall or sudden variation in temperature.

#### 5.4.3. Forecasting Uncertainty Using BNN and WBNN Models

[45] As described earlier, the BNN and WBNN models were developed using different realizations of the training data set, and the multiple forecasts obtained for the different realizations is used to assess the forecasting uncertainty associated with the water demand forecasts. A confidence band indicates the uncertainty associated with the forecasts; a narrow confidence band indicates less variability of the statistics with respect to possible future changes in the nature of the input data set and thus indicates that the model is robust [Khalil *et al.*, 2005].

##### 5.4.3.1. Uncertainty Assessment Using BNN Models

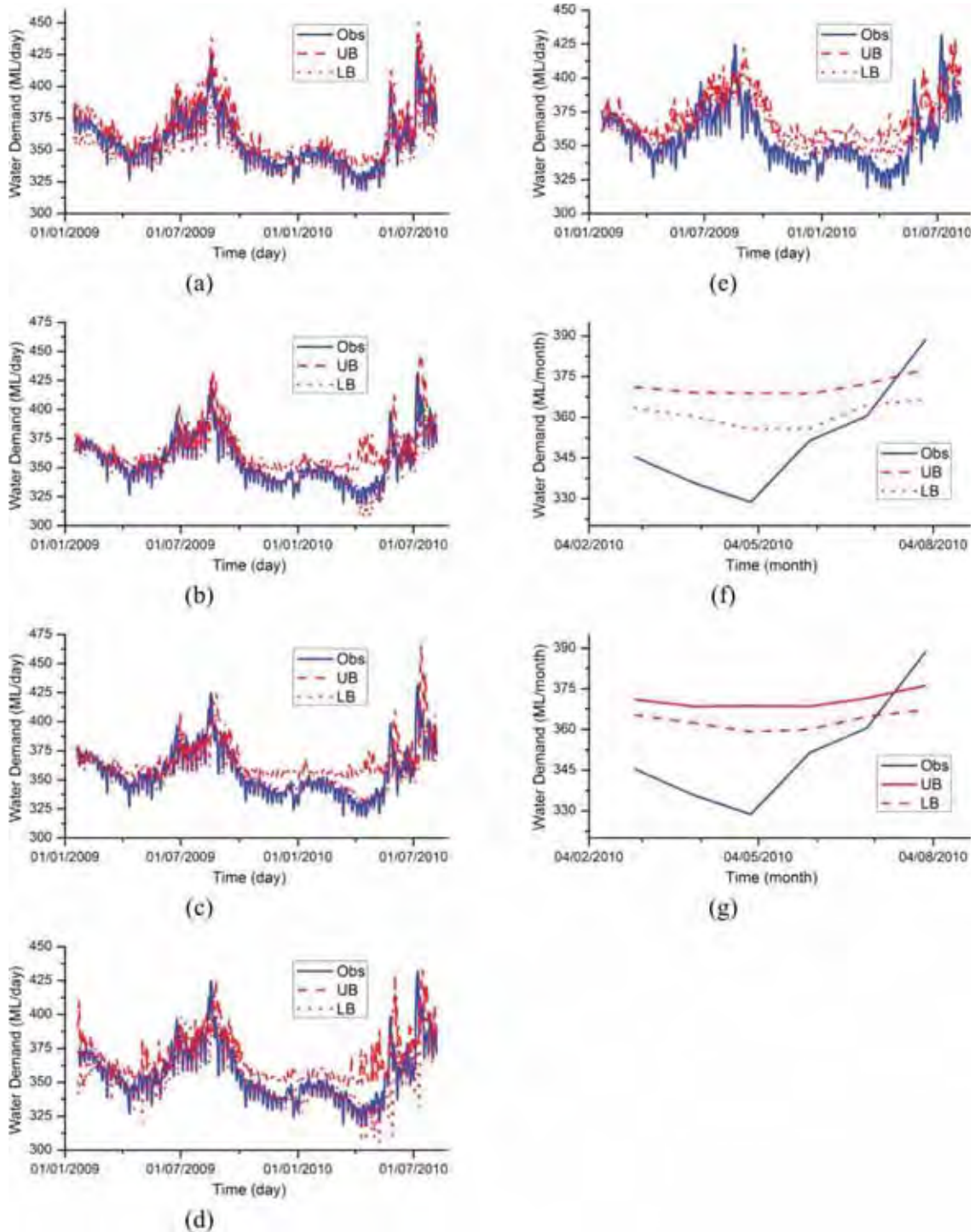
[46] The uncertainty associated with the NN model forecasts was quantified by building 95% confidence bands using BNN model forecasts. Wider confidence bands signify larger uncertainty and vice versa. Uncertainty associated with the daily, weekly, and monthly forecasts are shown in Figure 8 using BNN models. The figures show that as the lead time increases from 1 day to 5 days, 1 week to 2 weeks, and 1 month to 2 months, the uncertainty associated with the forecasts also increases. Further, it can be observed that there is more uncertainty during times of low water demand compared to during times of high water demand for a particular lead time for daily, weekly, and monthly forecasts. This phenomenon is more distinct for higher time steps or longer lead times. Moreover, several low water demand values fall outside the confidence band, especially for 3 and 5 day, 1 and 2 week, and 1 and 2 months lead time forecasts.

##### 5.4.3.2. Uncertainty Assessment Using WBNN Models

[47] Similar to the BNN models, the WBNN models were also used to assess the uncertainty associated with the WBNN forecasts by generating 95% confidence bands for 1, 3, and 5 day, 1 and 2 week, and 1 and 2 month lead time forecasts and are shown in Figure 9. The figures show that the forecasted confidence bands show the general behavior of the observed values. The WBNN forecasted confidence bands contain a higher number of observed water demand values in between the confidence bands compared to the BNN model forecasted confidence bands. This shows that WBNN models are more reliable compared to BNN models. Even though the width of the confidence bands is almost the same for 1, 3, and 5 day lead time water demand forecasts, it can be observed that the number of actual values included in the confidence bands decreases as the lead time increases. This phenomenon is more prominent for low water demand values. Comparing the performance with the BNN models it can be observed that the performance of the WBNN models is better for all the lead times, as the confidence bands better show the general behavior of the observed water demand values. As well, the number of actual values inside the confidence band is greater in WBNN predicted confidence bands compared to the actual values included in the BNN forecasted confidence bands.

[48] The actual number of values included in the confidence bands for 1, 3, and 5 day lead time forecasts using WBNN forecasted confidence bands is 94.9, 91.5, and 74.7% of the actual values for 1, 3, and 5 day lead time

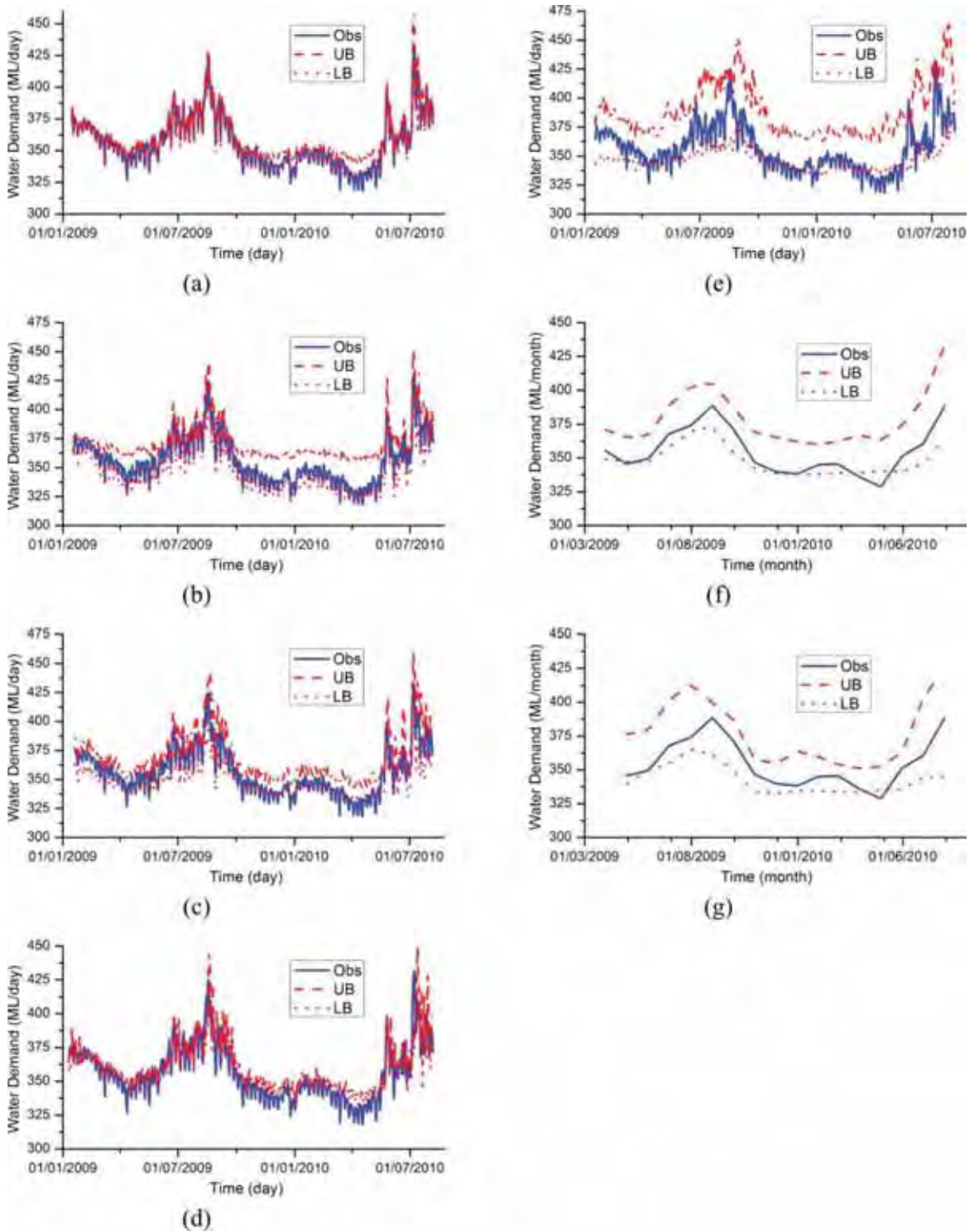




**Figure 8.** 95% confidence band with observed water demand in Montreal for (a) 1 day, (b) 3 day, (c) 5 day, (d) 1 week, (e) 2 week, (f) 1 month, and (g) 2 month lead time forecast using BNN models.

forecasts. For weekly water demand forecasts using WBNN models, the confidence bands are very narrow for 1 week lead time forecasts with only 17.6% of the actual or observed values included inside the confidence band, whereas in the case of a 2 week lead time, the confidence bands are very wide with 67.9% of the actual values

included inside the confidence bands. This shows that the WBNN model is capable of forecasting weekly water demand even for 2 week (i.e., 14 day) lead times, but the uncertainty associated with the forecasts can vary significantly. The 1 and 2 month lead time forecasted confidence bands using WBNN models show the general behavior of



**Figure 9.** 95% confidence band with observed water demand in Montreal for (a) 1 day, (b) 3 day, (c) 5 day, (d) 1 week, (e) 2 week, (f) 1 month, and (g) 2 month lead time forecast using WBNN models.

the average monthly water demand values. The forecasted confidence band is slightly wider for 2 month water demand forecasting compared to 1 month water demand forecasting, with 82.4% of the values actually included in the 95% confidence band for both 1 and 2 month lead

time average water demand forecasting. This indicates that the WBNN model is a suitable method for water demand forecasting and assessing uncertainty associated with the forecasted values for all the lead times explored in this study.

**5.4.4. Performance Comparison of WBNNs With Different Numbers of Bootstrap Resamples**

[49] As the performance of the WBNN models was found to be better than the BNN models for all the lead times (i.e., daily, weekly, and monthly), the performance of the WBNN model was also compared for different numbers of bootstrap samples (i.e., 25, 50, 200, and 500). In water resources studies, bootstrap resampling based modeling (BNN) has been applied in several studies, but the number of bootstrap resamples/realizations required for developing an accurate BNN model has not been explored to date. The performance of the WBNN model for 1, 3, and 5 day, 1 and 2 week, and 1 and 2 month lead time forecasts using different numbers of bootstrap resamples for Montreal is shown in Table 6. Even though it can be observed that for 1 day lead time forecasts, the performance of WBNN models with 25, 100, 200, and 500 bootstrap resamples is slightly better than the performance of the WBNN model developed using 50 bootstrap resamples, the actual number of values included in the confidence band forecasted using WBNN with 50 bootstrap resamples are higher. For 3 day lead time forecasts, the WBNN model with 100 bootstrap resamples performed better in terms of the different performance indices and also included a higher number of actual observed values in the confidence band. Similarly, for 5 day lead time forecasts, considering the performance indices and actual observed values included in the confidence band, the WBNN model developed using 100 bootstrap resamples performed better. For weekly water demand forecasts, 1 week ahead forecasts show better performance with 100 bootstrap resamples, whereas 2

week ahead forecasts are better with a higher number of bootstrap resamples (i.e., 200), and in the case of monthly forecasts, a lower number (i.e., 50 and 25 for 1 and 2 month lead time, respectively) of bootstrap resamples provides better model performance. It is clear from these observations that daily and weekly water demand forecasts, which are using daily time steps, required higher numbers of bootstrap resamples compared to monthly water demand forecasts. The reason is the higher correlation between the lagged variables, as is shown in Figure 1c, and the lower autocorrelation in the case of monthly water demand time series, as is shown in Figure 2c. Higher autocorrelation creates the problem of multicollinearity, and makes it difficult for the model to extract the optimum parameters. In addition, it is clear that 100 bootstrap resamples are appropriate for developing WBNN models with highly correlated data, whereas 25–50 bootstrap resamples yield good results with less correlated water demand time series data.

**6. Conclusions**

[50] Accurate and reliable urban water demand forecasting is necessary for effective and sustainable urban water resources planning and management. In this study, NN, WNN, BNN, WBNN, ARIMA, and ARIMAX models were developed for daily (1, 3, and 5 day), weekly (1 and 2 week), and monthly (1 and 2 month) lead time water demand forecasting for the city of Montreal in Canada. Through five performance indices consisting of the coefficient of determination ( $R^2$ ), root-mean-square error (RMSE), percentage deviation in peak ( $P_{dv}$ ), mean absolute

**Table 6.** Performance of WBNN Model for Water Demand Forecasting in Montreal for 1, 3, and 5 Day; 1 and 2 Week; and 1 and 2 Months Lead Time Forecasts using Different Numbers of Bootstrap Resamples

Performance Indices	Lead Time														
	Daily														
	1 day					3 day					5 day				
	No of Bootstrap Resamples														
	500	200	100	50	25	500	200	100	50	25	500	200	100	50	25
$R^2$	0.99	0.98	0.98	0.98	0.98	0.95	0.96	0.96	0.96	0.96	0.91	0.91	0.90	0.91	0.91
RMSE (ML/day)	4.08	4.24	4.09	4.43	3.82	6.64	6.16	6.08	6.57	6.40	8.13	7.75	7.29	7.49	7.48
$P_{dv}$ (%)	-1.07	-0.92	-0.97	-1.08	-1.72	0.90	1.05	1.04	1.1	1.32	1.09	0.91	0.54	0.26	-0.17
MAE (ML/day)	3.00	3.13	2.99	3.26	2.76	5.29	4.89	4.81	5.23	5.09	6.63	6.27	5.82	6.01	5.99
	Weekly														
	1 week					2 week									
$R^2$	0.88	0.74	0.78	0.77	0.77	0.63	0.59	0.68	0.56	0.58					
RMSE (ML/day)	10.31	12.69	10.06	11.85	11.77	18.58	16.05	18.76	19.86	20.13					
$P_{dv}$ (%)	-1.53	-5.97	-0.58	-1.95	-1.44	-0.21	0.54	1.47	1.67	0.15					
MAE (ML/day)	7.67	9.61	7.49	9.44	9.25	15.78	12.69	16.28	17.16	17.34					
	Monthly														
	1 month					2 month									
$R^2$	0.86	0.89	0.77	0.86	0.85	0.64	0.62	0.54	0.86	0.89					
RMSE (ML/month)	15.35	9.56	11.13	7.60	9.61	16.17	13.13	13.49	9.10	7.62					
$P_{dv}$	-4.39	-0.92	-0.68	0.56	-0.96	-0.80	-1.51	2.25	-0.63	-1.90					
MAE (ML/month)	13.92	7.48	8.62	6.23	7.87	13.09	10.43	10.94	7.30	6.80					



error (MAE), and persistence index (PI), it was found that the WNN and WBNN models performed considerably better than the NN, BNN, ARIMA, and ARIMAX models. These results attest to the ability of wavelet analysis to effectively decompose time series with nonstationary data into discrete wavelet components, allowing diagnosis of cyclic patterns and trends at varying temporal scales.

[51] It was found in this study that for longer lead times (i.e., weekly and monthly) where trend and nonstationarity is more pronounced, the performance of ARIMA and ARIMAX models is better compared to the simple NN and BNN models, as NN models were found to be weak in capturing these phenomena. However, wavelet analysis, which analyses the time series data in the time and frequency domain, helped extract the trend and nonstationarity from the data, and NN models developed using these extracted data (i.e., WNN and WBNN) improved significantly. Overall, it was found that the WBNN model provided significantly improved performance of daily, weekly, and monthly water demand forecasts and can be used to assess the uncertainty associated with the forecast to improve operational water demand forecasting.

[52] In this study, it was also found that the number of bootstrap resamples should not be taken as a default number; instead bootstrap-based forecasting models should be optimized carefully. For monthly lead times, bootstrap models developed using a smaller number of bootstrap resamples are appropriate, whereas for daily and weekly lead time forecasts a higher number of bootstrap samples can improve the performance of bootstrap-based neural network models (i.e., WBNN models). Uncertainty assessments were also performed in this study by assessing robustness through confidence bands, developed from the results of different realizations of BNN and WBNN model forecasts. In this study, confidence bands developed using WBNN models were better compared to confidence bands developed using BNN models. Overall, the use of wavelets improved the accuracy of the forecasts, while the use of bootstraps allowed for uncertainty testing and ensured model robustness along with improved reliability by reducing variance.

[53] Considering that the use of hybrid wavelet-bootstrap-neural network models is a new advancement in urban water demand forecasting, there are numerous areas for potential future work. Water demand varies between different days of the week [Herrera et al., 2010] as well as between day and night hours [Ghiassi et al., 2008], and this smaller-scale analysis could further improve forecasting and subsequent system operations. Additionally, the performance of different model hybrids with wavelets and bootstraps would be worth investigating, for example wavelet-bootstrap-support vector regression models. Moreover, Caiado [2010] has studied the possibility of combining forecasts derived from varying methods and data sets to improve accuracy, which would be an interesting possibility with WNN and WBNN models. Finally, the replicability of results from studies like this one in other cities would aid in better understanding and assessing these models.

[54] **Acknowledgments.** This research was funded by an NSERC Discovery grant, an FQRNT New Researcher grant, and a CFI grant held by Jan Adamowski.

## References

- Adamowski, J. F. (2008), Peak daily water demand forecast modeling using artificial neural networks, *J. Water Resour. Plann. Manage.*, 134(2), 119–128.
- Adamowski, J., and H. F. Chan (2011), A wavelet neural network conjunction model for groundwater level forecasting, *J. Hydrol.*, 407, 28–40.
- Adamowski, J., and K. Sun (2010), Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds, *J. Hydrol.*, 390, 85–91.
- Adamowski, J., H. Fung Chan, S. O. Prasher, B. Ozga-Zielinski, and A. Sliusarieva (2012), Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada, *Water Resour. Res.*, 48, W01528, doi:10.1029/2010WR009945.
- Altunkaynak, A., M. Özger, and M. Cakmakci (2005), Water consumption prediction of Istanbul City by using fuzzy logic approach, *Water Resour. Manage.*, 19, 641–654.
- Arhami, M., N. Kamali, M. M. Rajabi, (2013), Predicting hourly air pollutant levels using artificial neural networks coupled with uncertainty analysis by Monte Carlo simulations, *Environ. Sci. Pollut. Res.*, 20(7), 4777–4789, doi:10.1007/s11356-012-1451-6.
- Ascough, J. C., H. R. Maier, J. K. Ravalico, and M. W. Strudley (2008), Future research challenges for incorporation of uncertainty in environmental and ecological decision-making, *Ecol. Modell.*, 219, 383–399.
- Aubertin, L., A. Aubin, G. Pelletier, D. Curodeau, M. Osseyrane, and P. Lavallée (2002), Identifying and prioritizing infrastructure rehabilitation, *North Am. Soc. Trenchless Technol.*
- Babel, M. S., A. Gupta, and P. Pradhan (2007), A multivariate econometric approach for domestic water demand modeling: An application to Kathmandu, Nepal, *Water Resour. Manage.*, 21, 573–589.
- Barreto, H., and F. M. Howland (2006), *Introductory Econometrics: Using Monte Carlo Simulation with Microsoft Excel*, Cambridge Univ. Press, Cambridge.
- Bishop, C. M. (1995), *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, U. K.
- Caiado, J. (2010), Performance of combined double seasonal univariate time series models for forecasting water demand, *J. Hydrol. Eng.*, 15(3), 215–222.
- Campolo, M., P. Andreussi, and A. Soldati (1999), River flood forecasting with a neural network model, *Water Resour. Res.*, 35(4), 1191–1197.
- Cannas, B., A. Fanni, L. See, and G. Sias (2006), Data preprocessing for river flow forecasting using neural networks: Wavelet transforms and data partitioning, *Phys. Chem. Earth*, 31, 1164–1171.
- City of Montreal (2010), *The Montreal Community Sustainable Development Plan 2010–2015*, Ville de Montreal, Que.
- Efron, B. (1979), Bootstrap methods: Another look at the jackknife, *Ann. Stat.*, 7, 1–26.
- Efron, B., and R. J. Tibshirani (1993), *An Introduction to the Bootstrap*, Chapman and Hall, London, U. K.
- Environment Canada (2010), Available at <http://www.climate.weatheroffice.gc.ca>.
- Firat, M., M. E. Turan, and M. A. Yurdusev (2009), Comparative analysis of fuzzy inference systems for water consumption time series prediction, *J. Hydrol.*, 374, 235–241.
- Ghiassi, M., D. K. Zimbra, and H. Saidane (2008), Urban water demand forecasting with a dynamic artificial neural network model, *J. Water Resour. Plann. Manage.*, 134(2), 138–146.
- Haykin, S. (1999), *Neural Networks: A Comprehensive Foundation*, Prentice Hall, Englewood Cliffs, N. J.
- Heil, C. E., and D. F. Walnut (1989), Continuous and discrete wavelet transforms, *SIAM Rev.*, 31(4), 628–666.
- Herrera, M., L. Torgo, J. Izquierdo, and R. Pérez-García (2010), Predictive models for forecasting hourly urban water demand, *J. Hydrol.*, 387, 141–150.
- Hinsbergen, C., P. I. van, J. W. C. van Lint, and H. J. van Juylen (2009), Bayesian committee of neural networks to predict travel times with confidence intervals, *Trans. Res. Part C*, 17(5), 498–509.
- House-Peters, L. A., and H. Chang (2011), Urban water demand modeling: Review of concepts, methods, and organizing principles, *Water Resour. Res.*, 47, W05401, doi:10.1029/2010WR009624.
- Jain, A., and L. G. Ormsbee (2002), Short-term water demand forecast modeling techniques-conventional methods versus AI, *AWWA J.*, 94(7), 64–72.
- Jain, A., and A. M. Kumar (2007), Hybrid neural network models for hydrologic time series forecasting, *Appl. Soft Comput.*, 7, 585–592.

- Jia, Y., and T. B. Culver (2006), Bootstrapped artificial neural networks for synthetic flow generation with a small data sample, *J. Hydrol.*, 331, 580–590.
- Kame'enui, A. (2003), Water demand forecasting in the Puget Sound Region: Short and long-term models, MS thesis, Dep. of Civil and Environ. Eng., Univ. of Washington, Seattle, Washington.
- Kayaga, S., and I. Smout (2007), Water demand management: A key building block for integrated resource planning for the city of the future, First SWITCH Scientific Meeting, Univ. of Birmingham, U. K.
- Khalil, A., M. McKee, M. W. Kemblowski, T. Asefa, and L. Bastidas (2005), Multiobjective analysis of chaotic dynamic systems with sparse learning machines, *Adv. Water Resour.*, 29, 72–88.
- Kisi, O. (2010), Wavelet regression model for short-term streamflow forecasting, *J. Hydrol.*, 389, 344–353.
- Leclerca, M., and T. B. M. J. Ouarda (2007), Non-stationary regional flood frequency analysis at ungauged sites, *J. Hydrol.*, 343, 254–265.
- Lertpalangsunti, N., C. W. Chana, P. Mason, and P. Tontiwachwuthikul (1999), A toolset for construction of hybrid intelligent forecasting systems: Application for water demand prediction, *Artif. Intell. Eng.*, 13, 21–42.
- Maier, H. R., and G. C. Dandy (2010), Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications, *Environ. Modell. Software*, 15, 101–124.
- Maier, H. R., A. Jain, G. C. Dandy, and K. P. Sudheer (2010), Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions, *Environ. Modell. Software*, 25, 891–909.
- Mallat, S. G. (1989), A theory for multi resolution signal decomposition: The wavelet representation, *IEEE Trans. Pattern Anal. Machine Intell.*, 11(7), 674–693.
- Nourani, V., M. Komasi, and A. Mano (2009), A multivariate ANN-wavelet approach for rainfall–runoff modeling, *Water Resour. Manage.*, 23(14), 2877–2894.
- Rosso, O. A., A. Figliola, S. Blanco, and P. M. Jacovkis (2004), Signal separation with almost periodic components: A wavelets based method, *Rev. Mex. Fisica*, 50, 179–186.
- Sahoo, G. B., S. G. Schladow, and J. E. Reuter (2009), Forecasting stream water temperature using regression analysis, artificial neural network, and chaotic non-linear dynamic models, *J. Hydrol.*, 378, 325–342.
- Sharma, S. K., and K. N. Tiwari (2009), Bootstrap based artificial neural network (BANN) analysis for hierarchical prediction of monthly runoff in Upper Damodar Valley Catchment, *J. Hydrol.*, 374, 209–222.
- Srinivasulu, S., and A. Jain (2009), River flow prediction using an integrated approach, *J. Hydrol. Eng.*, 14, 75–83.
- Statistics Canada (2007), Available at <http://www12.statcan.ca>.
- Sudheer, K. P., A. K. Gosain, and K. S. Ramasastri (2002), A data-driven algorithm for constructing artificial neural network rainfall-runoff models, *Hydrol. Processes*, 16(6), 1325–1330.
- Tiwari, M. K., and C. Chatterjee (2009), Daily discharge forecasting using WANNs coupled with nonlinear bias correction techniques, *IAHS-AISH Publ.*, 331, 98–108.
- Tiwari, M. K., and C. Chatterjee (2010a), Uncertainty assessment and ensemble flood forecasting using bootstrap based artificial neural networks (BANNs), *J. Hydrol.*, 382, 20–33.
- Tiwari, M. K., and C. Chatterjee (2010b), Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN hybrid approach, *J. Hydrol.*, 394, 458–470.
- Tiwari, M. K., and C. Chatterjee (2011) A new wavelet-bootstrap-ANN hybrid model for daily discharge forecasting, *J. Hydroinf.*, 13(3), 500–519.
- Tiwari, M. K., K. Y. Song, C. Chatterjee, and M. M. Gupta (2013), Improving reliability of river flow forecasting using neural networks, wavelets and self-organizing maps, *J. Hydroinf.*, 15(2), 486–502, doi: 10.2166/hydro.2012.130.
- Twomey, J., and A. Smith (1998), Bias and variance of validation methods for function approximation neural networks under conditions of sparse data, *IEEE Trans. Syst. Man Cybernet. Part C: Appl. Rev.* 28(3), 417–430.
- Wu, C. L., K. W. Chau, and Y. S. Li (2009), Methods to improve neural network performance in daily flows prediction, *J. Hydrol.*, 372, 80–93.
- Wu, C. L., K. W. Chau, and C. Fan (2010), Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques, *J. Hydrol.*, 389, 146–167.
- Zhou, S. L., T. A. McMahon, A. Walton, and J. Lewis (2000), Forecasting daily urban water demand: A case study of Melbourne, *J. Hydrol.*, 36(3), 153–164.
- Zhou, S. L., T. A. McMahon, A. Walton, and J. Lewis (2002), Forecasting operational demand for an urban water supply zone, *J. Hydrol.*, 259, 189–202.





VOLUME 3 ISSUE 2

The International Journal of the  
**Constructed  
Environment**

 COMMON  
GROUND

[CONSTRUCTEDENVIRONMENT.COM](http://CONSTRUCTEDENVIRONMENT.COM)

# The International Journal of the Constructed Environment

.....  
VOLUME 3 ISSUE 2 2012



**THE INTERNATIONAL JOURNAL OF THE CONSTRUCTED ENVIRONMENT**  
www.constructedenvironment.com

First published in 2012-2013 in Champaign, Illinois, USA  
by Common Ground Publishing LLC  
www.commongroundpublishing.com

ISSN: 2154-8587

© 2012-2013 (individual papers), the author(s)  
© 2012-2013 (selection and editorial matter) Common Ground

All rights reserved. Apart from fair dealing for the purposes of study, research, criticism or review as permitted under the applicable copyright legislation, no part of this work may be reproduced by any process without written permission from the publisher. For permissions and other inquiries, please contact [cg-support@commongroundpublishing.com](mailto:cg-support@commongroundpublishing.com).

*The International Journal of the Constructed Environment* is peer-reviewed, supported by rigorous processes of criterion-referenced article ranking and qualitative commentary, ensuring that only intellectual work of the greatest substance and highest significance is published.

# Ecosystems Biomimetics: Ecological Systems Diagrams for Characterization of Environmental Performance of Buildings

Mercedes Garcia-Holguera, McGill University, Canada  
Grant Clark, McGill University, Canada  
Susan Gaskin, McGill University, Canada  
Aaron Sprecher, McGill University, Canada

*Abstract: Ecosystems biomimetics means learning from ecosystems; that is, learning from complex, resilient, self-organized systems in nature, and transferring valuable ecosystem patterns into the architectural work. However, the methodology is incipient and biomimetic tools have to cross disciplines to convey meaning for both ecology and architecture; so qualitative and quantitative tools need to be developed in order to stimulate the research in the field. In this paper, an ecological engineering tool, the energy systems diagrams as defined by ecologist H. Odum, is used to represent more than 20 sustainable-rated buildings under the light of ecological systems. The buildings selected are certified projects under the LEED, Living Building Challenge or Passive House rating systems, and data has been captured from their organizations web sites. The results show that ecological systems diagrams are powerful and effective instruments for characterization of the environmental performance of buildings in terms of energy and matter use; and that a shared language between disciplines is achievable. The validation of ecological systems diagrams as a useful biomimetic tool gives ground for further research on quantitative instruments to develop a complete methodology of ecosystems biomimetics.*

*Keywords: Sustainable Architecture, Biomimetics, Construction Ecology*

## Introduction

### *Biomimetics in Architecture*

**B**iomimetics, biomimesis, or biomimicry means learning from nature to solve human problems (Benyus 1997). It also has been described as the transfer of knowledge from biology towards the engineering fields (Gruber 2011), and another definition describes biomimetics as the adaptation of mechanisms and functions of biological sciences in engineering, design and other disciplines (DTI 2007). Examples of human technology copying nature can be found throughout history. Famous examples include Leonardo da Vinci's airplane drawings, and the invention of Velcro™ by George de Mestral through the observation of cockleburs attached to his dog's fur. In architecture, the works of John Paxton (Crystal Palace, London, UK), Antoni Gaudi (e.g. Sagrada Familia, Barcelona, Spain) or Frei Otto (Munich Olympic Stadium, Munich, Germany) are in some aspects related to the biomimetic process; however, only during the last decade has architectural biomimetics been the subject of rigorous research. This research has been motivated by technological and computational advances, as well as by successful biomimetic results in other disciplines. A more recent example of biomimetics in architecture is the design of the Eastgate Building in Harare by Mick Pierce (built in 1996), which mimics a termite mound to optimize heating and cooling.

A commonly-accepted defining feature of biomimetics is transdisciplinarity. In biomimetic architecture, transdisciplinarity means that biologists, ecologists, engineers and designers collaborate in order to generate knowledge. Nevertheless, this is a difficult task and much effort is placed on defining the biomimetic approach and its methods for transferring natural patterns into buildings.

### *A Biomimetic Classification for Architecture*

Pedersen (2007) has created a framework to classify architectural biomimetics according to the phenomena in nature that are the source of inspiration: organisms, organisms' behavior, and the functioning of ecosystems. Organism biomimetics refers to the study of one organism or one part of an organism (e.g. lotus leaves for Lotusan™ paint); behavioral biomimetics means that one function of an organism or species is examined in the context of its surroundings; finally, ecosystem biomimetics involves mimicking a group of functions and processes that relate biotic and abiotic components. This document focuses on this latter level of biomimetics that has also been described as ecomimicry, ecomimesis or ecomimetics.

Some benefits of ecosystem biomimetics are that it can encompass organism and behavior biomimetics; it accommodates the combination of biomimetic strategies with other building sustainable methods; and it can be applied to a wide variety of structures (from residential houses to urban zones). Finally and more importantly, ecosystem biomimetics has the potential to positively affect the environmental performance of buildings (Pedersen 2007). Some characteristics of certain ecosystems that could be mimicked in the building environment include:

- Effective use of solar energy;
- Thermodynamic efficiency
- Complexity
- Informational richness
- Adaptability

However, biomimetics is not necessarily equivalent to sustainability or an excellent environmental performance of buildings. A given biomimetic design might be relatively unsustainable compared to other alternatives; for example aircrafts, whose design was originally inspired by birds, but consume huge amounts of fossil fuels and pollute the atmosphere. To develop an ecosystem biomimetic approach that is also environmentally oriented (i.e. results in less use of non-renewable resources and less pollution), it is necessary to explicitly address the issue. Some architectural parameters related to the environmental performance of buildings are: water consumption, energy use, greenhouse gas (GHG) emissions, indoor air quality, waste management, orientation and insulation of the buildings, materials and user behavior. These parameters will be considered in the development of an ecosystem biomimetic methodology.

### **Problem: Biomimetic Methodology**

A complete and exhaustive methodology on architectural biomimetics does not exist. It is widely accepted, however, that the biomimetic process can be approached from two different perspectives: the top-down perspective and the bottom-up perspective (DTI 2007; Pedersen 2007; Helms, Vattam, and Goel 2009; Gruber 2011). In the first case, a design or engineering problem is described and a solution is sought in the natural realm. In the bottom-up approach a professional from the natural science disciplines identifies an interesting phenomenon in nature and applications are sought in human technology. Case studies of each of these approaches are presented in the literature. Helms et al. (2009), for example, describe a number of steps from problem-definition to principle-application, and Gruber (2011) reports cases from research and student projects. Despite the acceptance of the top-down and bottom-up approaches, the lack of qualitative and quantitative tools impedes progress towards an effective biomimetic methodology. However, three tools are identified in the report from the Department of Trade and Industry in the UK (DTI 2007). First Bio-TRIZ, which is a problem-solving technology based on TRIZ (Theory of Inventive Problem Solving) that works with an extensive database comprising over three million patents (BioTRIZ, 2012); second, a complementary database at the Max Planck Institute; and finally a lexicon research method. None of these tools are specifically



focused on ecosystem biomimetics. The present document addresses this point and presents a qualitative tool adapted from ecological engineering as a common language for the characterization of ecosystems and architectural systems. The document is organized as follows; first, there is an introduction to energy systems diagrams as defined by Odum (1994); second, the correlation between architectural components and ecological components is detailed; and finally several case studies are analyzed to evaluate the appropriateness of this tool as a common conceptual platform for knowledge transfer between architecture and ecology.

## **Energy Systems Diagrams from Ecological Engineering**

### ***Ecological Engineering***

The term “ecological engineering” was first coined by Howard T. Odum in the early 1960’s (Mitsch and Jørgensen 2003), and he defined ecological engineering as the “environmental manipulation by man using small amounts of supplementary energy to control systems in which the main energy drives are still coming from natural sources” (Mitsch and Jørgensen 2003). Afterwards, Mitsch and Jørgensen (2003) defined ecological engineering as “the design of sustainable ecosystems that integrate human society with its natural environment for the benefit of both.” This involves the restoration of damaged ecosystems and the development of new ecosystems in order to solve environmental problems. Included in these problems are natural resource depletion and excessive GHG emissions. Furthermore Mitsch and Jørgensen (2003) listed five essential characteristics of ecological engineering: “(1) it is based on the self-designing capacity of ecosystems; (2) It can be the acid test of ecological theories; (3) It relies on system approaches; (4) It conserves non-renewable energy sources; and (5) It supports biological conservation”.

The third characteristic, the system approach, is essential for this work. The word ‘ecosystem’ is an abbreviation of ‘ecological system’: “an organized system of land, water, mineral cycles, living organisms, and their programmatic behavioural control mechanisms” (Odum 1994). A system is a set of components interacting with one another. Von Bertalanffy (2008) states that “there are principles which apply to systems in general, whatever the nature of their component elements or the relations or ‘forces’ between them”; and Odum applied this holistic approach from systems theory to ecology in order to develop his energy systems diagrams.

### ***Energy Systems Diagrams***

The energy systems diagram is a “methodology for converting verbal models into system network diagrams showing mathematic, energetic, cybernetic and hierarchical attributes simultaneously for many purposes” (Odum and Peterson 1996); that is, energy systems diagrams are a language that represents the organization of systems and their flows of energy. This language can be used to characterize environmental systems as well as other social or human systems.

When Odum (1994) began representing flows of energy in ecological systems, he turned toward the electrical engineering language, and found that the symbols and signs used in electronic circuits were able to represent organization and energy dynamics in ecosystems. Before long the energy systems diagrams “grew out of recognition and appreciation for open system thermodynamics of ecosystems, general systems theory, and simulation” (Brown 2004). The energy systems diagrams can be explained through their elements and the different steps for building them.

The elements of an energy systems diagram can be split into: sources, state variables, energy flows, and the boundary. The sources are the external sources of energy. The state variables, or components, represent the elements in the system that store and/or transform energy. Energy






might be represented as materials, information or 'pure energy' (Odum 1994). The energy flows, or pathways, represent the energy per unit time; and flows have no storage capacity. The boundary defines the limits of the system under study, and separates the components from the external sources or driving forces. Odum selected different symbols to represent these elements and their variations. Table 1 shows the qualitative descriptions of the symbols used in this document. Brown (2004) describes the process of diagramming as a three-stage process: (1) identify the external sources of energy and define the boundary of the system; (2) draw the components; and (3) describe the outputs. Odum placed the elements in the diagram from left to right according to the increase in their energy quality. When Odum defined higher energy quality he was referring to that energy "that is more concentrated and in a form capable of special actions when fed back" (Odum, 1994). This energy quality is measured by the amount of solar calories needed to generate one calorie of another type of energy; so, for example, the source representing sun's energy in a diagram will be on the left side of the source representing fossil fuels energy.

### **Equivalences between Systems Diagrams and Architectural Structure**

The energy systems diagrams can be used to represent the flows of energy in many different systems. Mainly, they have been used to represent environmental systems; however, other social or economic systems can be represented using the diagrams. Odum, for example, applied them to explain the "functions and relationships of a familiar system" (Odum 2007) and also to characterize a city (Odum 1994). More recently, a group of researchers used energy systems diagrams to represent a building in the framework of energy (with an m) analysis (Srinivasan et al. 2012). In the present document the energy systems diagrams are used in order to validate two ideas. First, that energy systems diagrams can represent the environmental performance of buildings. Second, that energy systems diagrams can be used as a platform for transferring patterns from ecological systems towards architectural systems. In the aforementioned examples by Odum (1994; 2007) and Srinivasan (2012) it is clear that energy systems diagrams can represent certain characteristics of buildings. This document will show that the energy systems diagrams can also represent parameters commonly used to determine the environmental performance of buildings (e.g., water consumption, energy use intensity, air quality), and that the sum of these parameters gives a holistic image of the environmental performance of a building.

The process of representing buildings using the energy systems diagrams requires a correlation between the symbols defined by Odum (1994) for the ecological engineering field and the building components. Table 1 presents the symbols used in this work and their equivalencies with architectural elements.

Table 1: Correlation between Energy System Symbols and Their Definitions in Ecological Engineering and Architecture.

Symbol	Ecological Engineering		Architecture	
	Definition	Examples	Definition	Examples
	"Outside source of energy delivering forces according to a program controlled from outside."	Sun Water Wind	Outside source of energy delivering forces according to the functional needs of the building.	Fossil Fuels Municipal Water Materials Information Natural sources (e.g., sunlight, wind, rain)
	"A compartment of energy storage within the system storing a quantity as the balance of inflows and outflows; a state variable"	Nutrients Detritus	Unit for energy storage in the building. It might be 'pure energy', materials or information.	Building envelope Water tanks Green roofs
	"Unit that collects and transforms low-quality energy under control interactions of high-quality flows."	Shore Plants Phytoplankton	Unit that collects and transforms low-quality energy under control interactions of high-quality flows of energy. The results are higher-quality energy flows necessary for the functioning of the building.	Photovoltaic panels Thermal Panels Wind turbines
	"Interactive intersection of two pathways coupled to produce and outflow in proportion to a function of both"	Chemical reactions	Unit that receives two energy flows and transforms them into a proportional energy flow of both.	Inverter Solar pump Heat pump Fog catcher
	"Unit that transforms energy quality, stores it, and feeds it back autocatalytically to improve inflow"	Fish Zooplankton Bacteria	Unit that transforms energy quality and can also store it and use it in order to maximize the inflows of energy. The functioning of these units is not based on efficiency, but rather on maximum useful power.	Building systems (hydraulic, lighting, hot water, heating, electrical, etc.) Garden Pond People

Note: Description of symbols used in energy system diagrams as defined by Odum (1994). The examples accompanying the ecological engineering definition of the symbols refer to an aquatic ecosystem (ibid). This is not an exhaustive table of examples and symbols.

## Representation of Buildings Using the Energy Systems Diagrams

### *Criteria for Case Studies Selection and Description of Diagramming Process*

Two basic criteria were followed to select the buildings represented using the energy systems diagrams: (1) buildings needed to embrace environmental measures in terms of their design and performance; (2) enough data should be available for the representation of the buildings and also for the analysis of the results.

As a result of these criteria a total of 29 buildings were selected from the U.S. Green Building Council (USGBC, 2012) database, the Living Building Challenge database (International Living Future Institute, 2012), and the Passive House database (Passive House Institutions, 2012). The 29 buildings have had included, at some degree, environmental principles in their design. In addition, the above mentioned organizations have public databases with enough data to develop this exercise. Lastly of the rating systems used to compile these lists, the Passive House rating system offers an approach focused on energy consumption reduction, while the other two rating systems emphasize additionally other environmental parameters such as water consumption, indoor air quality or materials selection.






The buildings selected from the USGBC database were taken from those rated under the LEED for New Construction v2.1 and LEED for New Construction v2.2. The main reason is that

the New Construction classification comprises all types of buildings (e.g. residential, offices, commercial, educational) and it provides complete information on all the building components. Versions 2.1 and 2.2 provide more recent data. Data was accessed between June-September 2012.

Once the 29 buildings were selected, an exhaustive review of the available data was conducted before the diagramming phase started. First, the boundaries of the building systems were defined: an external boundary representing the building site boundary; and an internal boundary representing the building envelope. Then, the sources or external driving forces were identified, followed by the description of the components of each building. The sources and the components were arranged in order of energy quality from left to right (Odum 1994). Flows of energy and component interactions were then indicated. Where data were unavailable, two options were offered. The first option was an assumption of interactions among components when that interaction was essential for the functioning of the building (e.g. there is no data for water consumption, but the building has a hydraulic system; therefore, water inflow has to be represented). The second option was to ignore the element or interaction in the diagram when data were unavailable; this was more common (e.g., if no data suggested the presence of control equipment, then that unit would not be represented in the diagram).

Energy flows were categorized as per Table 2.

Table 2: Description of Energy Flows Represented in the Energy System Diagrams.

<b>Energy Flows</b>		
<b>Symbol</b>	<b>Name</b>	<b>Definition</b>
	Inflows	Flow of energy entering the system or a component in the system
	Outflows	Flow of energy exiting the system.
	Control flows	Feedback flows of information that control the flows of a stock
	Recycle flows	Flows that represent the energy that is reused within the system
	Heat sink	Flows of energy dispersed as heat

***Energy Systems Diagrams Representation***

Energy systems diagrams of 29 buildings were developed. In this document four diagrams are shown at full scale. All remaining diagrams can be provided if requested, but are omitted in this document to facilitate the reading. The diagrams selected are representative of each of the rating systems, and display a wide variety of design strategies.

Figure 1: Great River Energy Building (LEED)

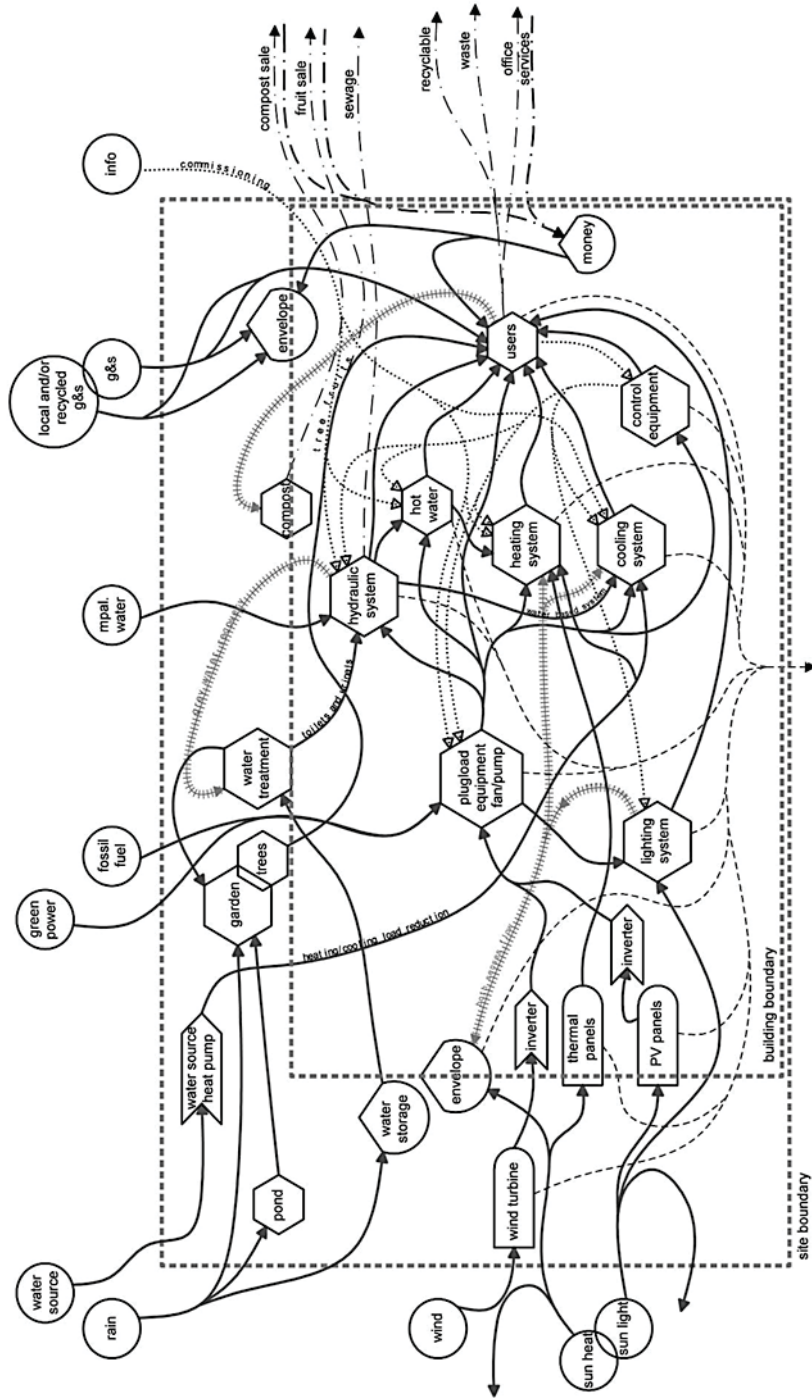




Figure 2: Chartwell School (LEED)

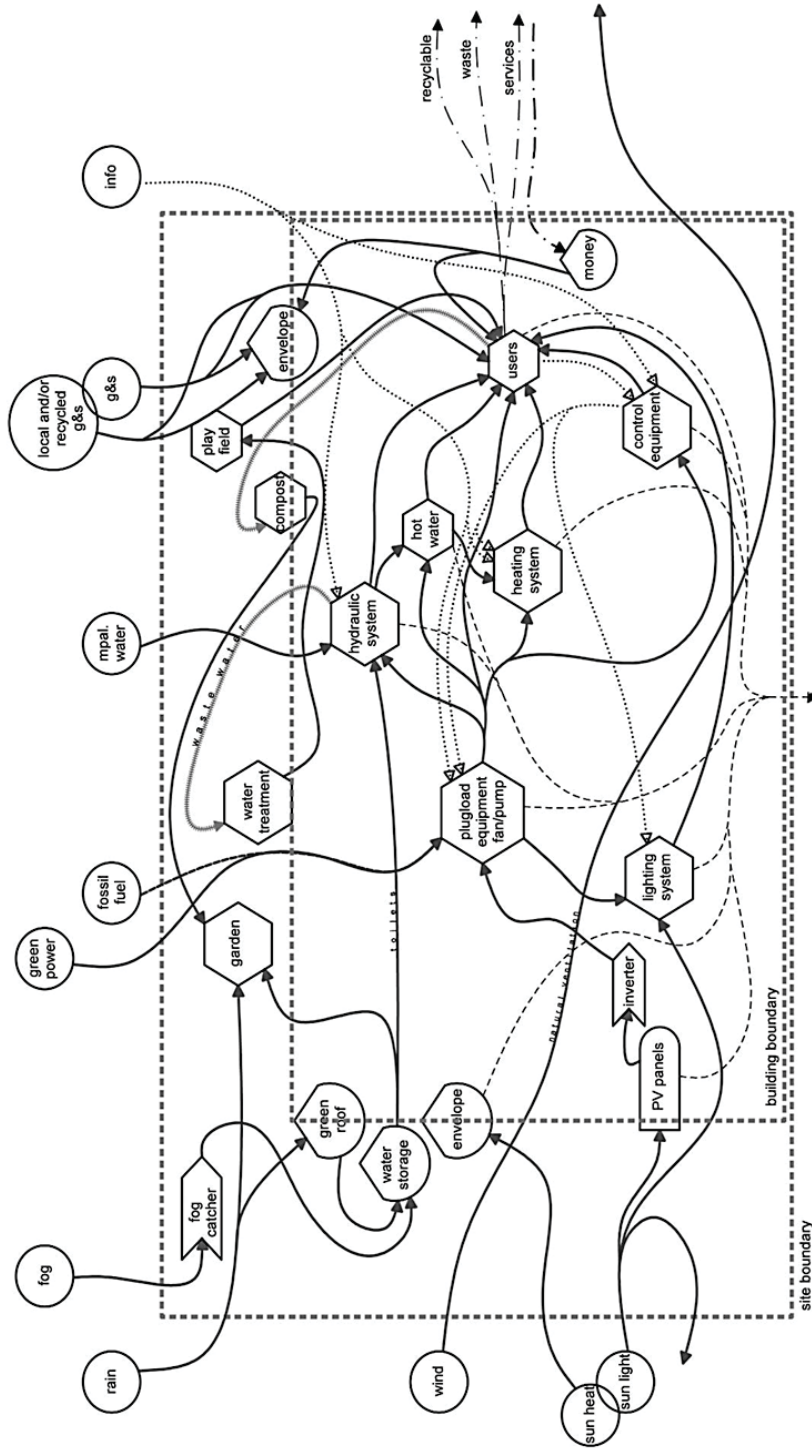


Figure 3: Omega Center (Living Building Challenge)

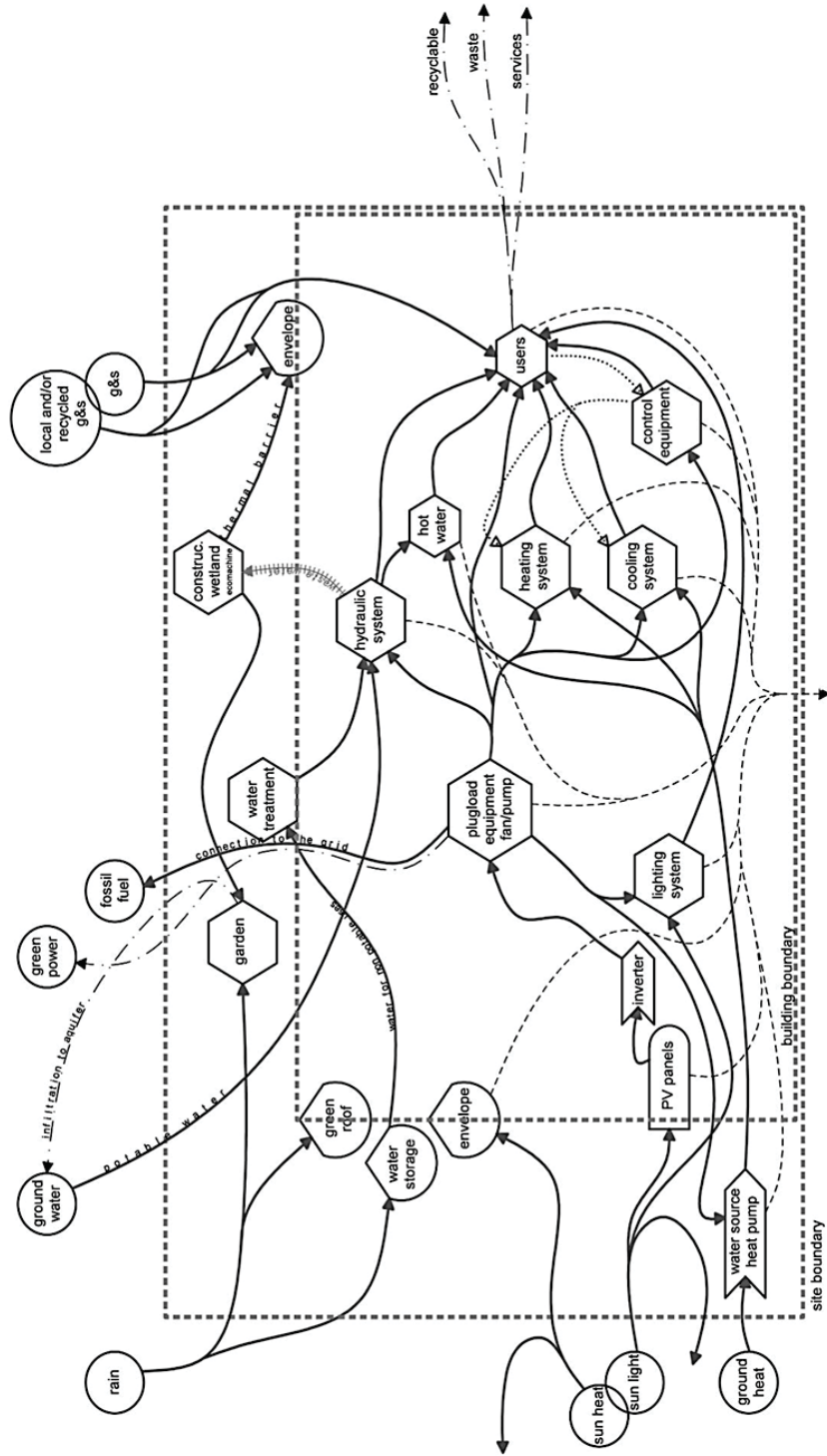
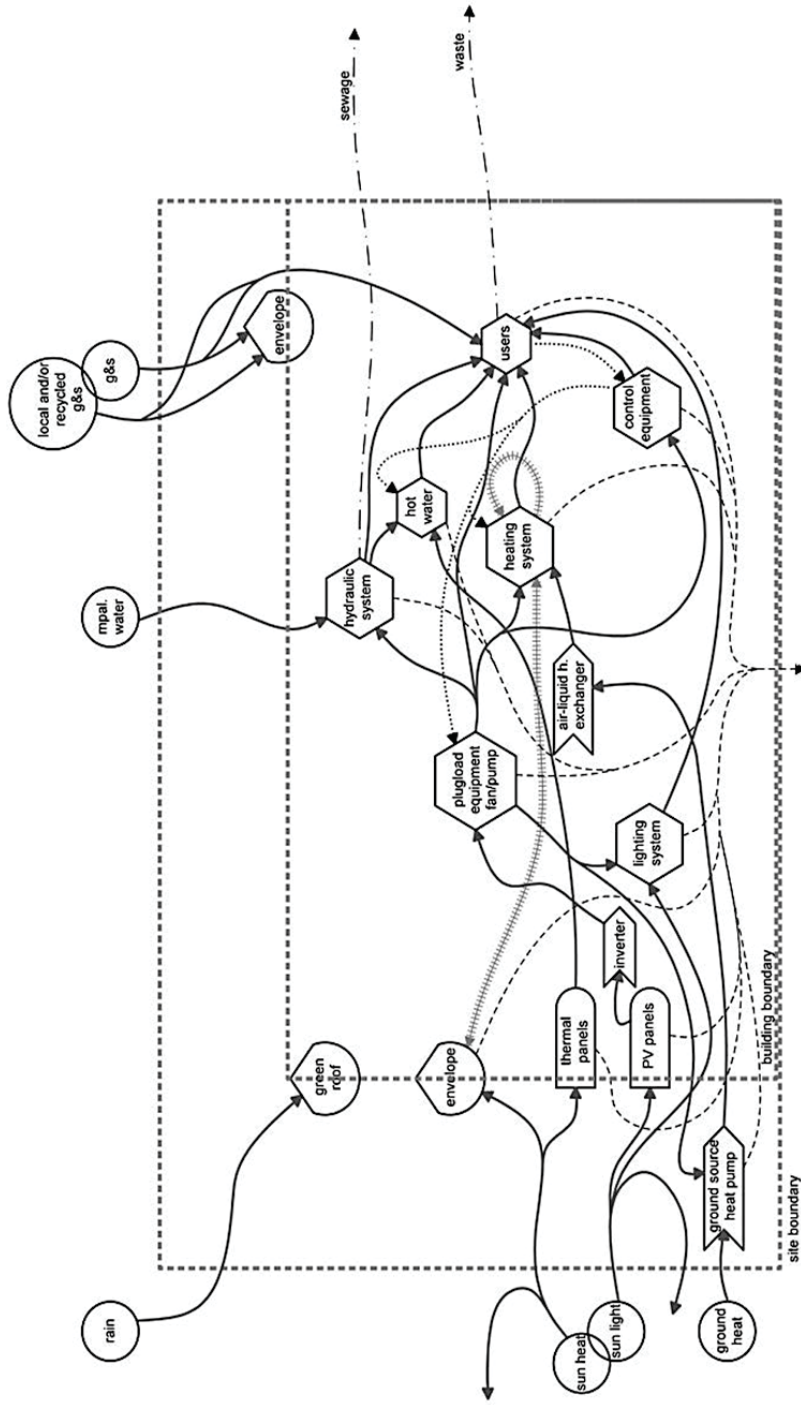


Figure 4: Family House in Hudson (Passive House)



### *Analysis of the Energy Systems Diagrams*

Each of the buildings shows several design strategies oriented towards environmental performance optimization. The energy systems diagrams are able to represent these design strategies and the multiple interactions that these strategies sometimes involve. For instance, the Chartwell School diagram (Figure 2) indicates rain water entering the building system and interacting with at least seven different components of the building (green roof, garden, water storage, hydraulic system, users of the building, water treatment plant, and playing field). In addition, the rainwater goes through several processes and transformations (storage, filtering, disposal, recycling, and reuse). The Chartwell School diagram also represents all the other environmental strategies implemented in the building in a holistic way; and it is able to characterize the diverse interactions among the components in a manner that is easily understood.

In addition to drawing the energy systems diagrams, an analysis of the parameters was carried out. The purpose of this analysis was to show the relationship between architectural parameters and energy systems diagrams parameters. Several parameters related to the environmental performance of buildings were selected: annual water consumption ( $l\ m^{-2}\ yr^{-2}$ ), energy use intensity ( $kwh\ m^{-2}\ yr^{-2}$ ), annual energy breakdown ( $kwh\ m^{-2}\ yr^{-2}$ ), annual percentage of renewable energy produced on-site ( $kwh\ m^{-2}\ yr^{-2}$ ), rating system, certification level, and building size ( $m^2$ ). Parameters associated with the energy system diagrams were chosen, for instance: type and number of components, components' interactions (connectivity), energy flows, and energy quality ratio. Table 3-A presents the available data on architectural parameters of environmental performance from the rating system's websites (USGBC, Living Building Challenge and Passive House); while table 3-B presents the parameters from the energy systems diagrams.

Table 3-A: Buildings' Parameters as per Data from the Rating Systems' Databases (LEED, Living Building Challenge, Passive House)

Buildings' Performance	Building Type	AWC l/m <sup>2</sup>	EUI* Kwh/m <sup>2</sup>	Total Annual Energy Breakdown (Kwh/m <sup>2</sup> )											TARE POS (Kwh/m <sup>2</sup> )			
				H	%	C	%	L	%	P/E/F	%	HW	%	O	%	Kwh/m <sup>2</sup>	%	
<b>LEED BUILDINGS NC v2.2</b>																		
PRESIDENT LINCOLN'S COTTAGE	Museum	337.00	293.62	29.36	10.00	44.32	15.09	116.34	39.62	96.67	32.92				6.93	2.36	0	0.00
NEVA SCHOOL	Education	164	74.20	20.27	27.32	0.87	1.17	22.49	30.31	16.14	21.75	14.43	19.45				2.3	31.00
ATLANTIC FLEET DRILL HALL	Military	379.00															0	0.00
SIGLER OFFICE AND WAREHOUSE	Office + Warehouse	129	77.00	12.80	16.62	7.39	9.60	19.94	25.90	32.84	42.65	4.09	5.31				0	0.00
GREAT RIVER ENERGY BUILDING	Office	201.93	28.80	14.26	19.52	9.67	30.07	14.89	99.90	49.47	4.51	2.23	15.09	7.47	31	15.35	0	0.00
IFAW WORLD HEADQUARTERS	Office	213.84	66.75	31.21	15.29	7.15			99.74	46.64	5.62	2.63					0	0.00
YALE SCHOOL OF FORESTRY AND ENV.	Education	83.93	6.86	8.17	6.06	7.22			0.0049	0.01				70.91	84.49	19.05	22.70	0.00
MATAROZZI-PELSINGER OFFICES	Office		58.44														20.88	35.73
KAUST CAMPU'S	Education		251														19.33	7.70
<b>LEED BUILDINGS NC v2.1</b>																		
BOSTON CHILDREN'S MUSEUM	Museum	173	562	329.63	58.65	41.27	7.34										0	0.00
ORNL JOINT INSTITUTE FOR COMPUTATIONAL SCIENCE	Education	260	479.21														0	0.00
GISH APARTMENTS	Residential	259	56.78	10.6	18.67	4.59	8.08	8.22	14.48	16.78	29.55	16.48	29.02				6.08	10.71
AIR FORCE WEATHER AGENCY HEADQUARTERS	Military		238.16	45.98	19.31	39.88	16.75	25.7	10.79	126.58	53.15						0	0.00
0142 CNT RENOVATION	Offices	275	178.94	61.29	34.25	19.77	11.05			93.9	52.48	4.23	2.36				9	5.03
JEWISH RECONSTRUCTION CONGREGATION	Religious		157.33														0	0.00
SHANGRI LA BOTANICAL GARDENS	Museum		56.5	0.19	0.34	37.11	65.68	10.74	19.01	24.93	44.12	0.71	1.26				6.64	11.75
CHARTWELL SCHOOL	Education	3810	85.59	40.44	47.25	0	0.00	11.96	13.97	15.04	17.57	11.19	13.07				12.35	14.43
<b>LIVING BUILDING CHALLENGE</b>																		
HAWAII PREPARATORY ACADEMY	Education	34	34.75	0			0	5.71	16.43	29.03	83.54						71.15	204.75
OMEGA CENTER	Education	125	41.35														67.23	162.39
TYSON LIVING LEARNING CENTER	Education	178	77.42														83.58	107.96
ECO SENSE	Residential	340	107.75														10.64	9.87
PAINTERS HALL	Office		65.71	6.44	9.80	6.44	9.80	17.07	25.98	34.64	32.72	1.16	1.77				73.31	111.57
IDEAS 22 DESIGN FACILITY	Office		66.68														68.44	102.64
<b>PASSIVE HOUSE</b>																		
FAMILY HOUSE HUDSON, US	Residential		66	12	18.18												66	100.00
RURAL REGENERATION CENTER, U	Education		86	15	17.44												0	0.00
STANDINGS COURT RESIDENCES, U	Residential		98	14	14.29												0	0.00
DAS BIOHAUS, US	Education		83	14	16.87												0	0.00

AWC: Annual Water Consumption; EUI: Energy Use Intensity; H: Heating; C: Cooling; L: Lighting; P/E/F: Plug-Loads/Equipments/Fans-Pumps; HW: Hot Water; O: Others; TARE POS: Total Annual Renewable Energy Production on Site

\*In terms of Site Energy for LEED and Building Challenge. In terms of Primary energy for Passive House For Passive House Buildings. Calculation of Site Energy equivalent: for US buildings Assumption: source energy electricity and a Primary Energy factor of 3.34 according to the ENERGY STARS Performance Ratings ([http://www.energystar.gov/about/business/evaluate\\_performance/site\\_source.pdf#8ae2-9c3a](http://www.energystar.gov/about/business/evaluate_performance/site_source.pdf#8ae2-9c3a)). For UK buildings: Assumption: source energy electricity and a Primary Energy factor of 2.58 according to SAP 2012 from the UK ([www.bre.co.uk](http://www.bre.co.uk))



Table 3-B: Diagrams' Parameters as per Data Obtained from the Energy System Diagrams

Diagrams' Parameters	S			NoCp			Stocks			Flows																																				
	R			BB			SB			Pr			St			Cm			Inflows				Outflows				R-P		C-P		Energy Quality															
	n	r	r	b	b	b	s	s	s	p	p	p	s	s	s	t	t	t	c	c	c	e1	e2	e3	e4	e5	e6	e7	e1	e2	e3	e4	e5	e6	e7	e1	e2	e3	e4	e5	e6	e7				
<b>LEED BUILDINGS NC v2.2</b>	3	5	9	1	0	1	8	15	4	9	4	1	3	4	3	3	5	6	4	9	2	3	8	5	3	2	4	5	7	1	3	2	3	3	4	5	7	1	3	2	3	3	4	5		
PRESIDENT LINCOLN'S COTTAGE	5	4	12	1	1	2	9	17	7	10	4	1	3	4	3	4	3	5	4	9	2	3	8	5	3	2	4	4	5	1	4	3	4	6	5	7	1	4	3	4	6	5	7			
NUEVA SCHOOL	3	5	9	3	0	2	10	16	5	9	3	0	3	4	3	3	4	3	9	2	3	8	4	3	2	4	3	5	1	3	2	3	3	4	5	7	1	3	2	3	3	4	5			
ATLANTIC FLEET DRILL HALL	3	5	10	3	0	2	11	11	4	7	4	0	4	4	3	2	4	4	3	10	2	3	9	5	3	4	4	7	1	3	2	3	3	4	5	7	1	3	2	3	3	4	5			
SIGLE OFFICE AND WAREHOUSE	6	5	14	7	3	2	13	20	10	10	6	3	3	4	3	5	7	4	10	2	3	8	6	3	3	5	5	7	1	5	4	5	7	6	8	0	1	3	2	3	5	4	6			
GREAT RIVER ENERGY BUILDING	4	4	8	2	0	1	9	10	4	6	3	0	3	4	2	1	2	3	2	7	2	3	8	5	3	3	4	0	1	3	2	3	3	4	0	1	3	2	3	5	4	6				
IFAW WORLD HEADQUARTERS	6	3	13	3	2	2	10	14	9	5	3	0	3	4	3	4	5	3	9	2	3	10	4	3	2	4	3	7	1	5	4	7	6	6	7	0	1	3	2	3	5	4	6			
YALE SCHOOL OF FORESTRY AND ENV.	4	5	11	1	1	2	7	18	7	11	3	0	3	4	3	3	4	4	3	9	2	3	8	3	3	4	3	4	1	5	4	5	7	6	8	0	1	3	2	3	5	4	6			
MATAROZZI-PELSINGER OFFICES	5	4	12	2	2	1	9	16	10	6	2	0	2	3	3	3	4	4	8	1	2	8	5	2	4	1	6	1	5	4	5	7	6	8	0	1	3	2	3	5	4	6				
KAUST CAMPUS	4	5	9	4	0	3	10	17	6	11	3	0	3	5	3	3	4	4	8	3	3	8	3	3	3	4	4	5	1	3	2	3	3	4	5	1	3	2	3	3	4	5				
<b>LEED BUILDINGS NC v2.1</b>	3	5	10	1	0	1	10	14	3	11	4	1	3	4	3	3	3	6	4	9	2	3	8	4	4	5	6	2	1	4	3	4	6	5	7	1	3	4	6	5	7					
BOSTON CHILDREN'S MUSEUM	5	4	11	2	1	1	12	6	6	6	3	1	2	4	3	2	3	3	2	9	2	3	9	4	2	5	2	3	1	5	4	6	6	7	0	1	3	2	3	5	4	6				
ORNL JOINT INSTITUTE FOR COMPUTATIONAL SCIENCE	3	5	9	1	0	1	9	9	4	5	4	1	3	3	3	2	2	9	1	2	8	4	2	2	5	0	6	1	3	2	3	3	4	5	7	1	3	2	3	3	4	5				
AIR FORCE WEATHER AGENCY HEADQUARTERS	5	4	12	2	1	1	11	13	7	6	3	1	2	3	3	4	2	2	9	1	2	8	5	2	4	0	6	1	5	4	5	6	7	0	1	3	2	3	5	4	6					
0142 CNT RENOVATION	4	5	10	1	0	2	9	11	4	7	3	0	3	4	3	2	4	4	2	9	2	8	4	3	3	4	1	6	1	3	2	3	3	4	1	6	1	3	2	3	5	4	6			
JEWISH RECONSTRUCTIONIST CONGREGATION	5	5	12	6	1	2	10	22	10	12	4	2	2	3	3	3	5	4	3	10	1	2	9	5	2	5	0	5	1	5	4	6	7	8	0	1	3	2	3	5	4	6				
SHANGRI LA BOTANICAL GARDENS	6	5	13	5	1	3	11	18	8	10	3	1	2	3	3	5	4	4	2	9	1	2	7	4	2	6	2	4	1	5	4	6	7	8	0	1	3	2	3	5	4	6				
CHARTWELL SCHOOL	5	2	16	5	2	5	11	13	9	4	5	3	2	3	3	2	2	3	2	9	1	2	9	4	2	5	1	4	1	5	4	5	7	6	8	0	1	3	2	3	5	4	6			
<b>LIVING BUILDING CHALLENGE</b>	5	2	12	5	1	3	11	12	8	4	5	3	2	3	2	1	3	2	2	9	1	2	9	4	2	4	1	1	1	5	4	6	7	8	0	1	3	2	3	5	4	6				
HAWAII PREPARATORY ACADEMY	4	2	12	4	1	2	12	10	6	4	5	3	2	4	3	1	2	4	2	9	2	2	9	5	4	2	8	4	1	5	4	5	6	7	0	1	3	2	3	5	4	6				
OMEGA CENTER	6	3	13	2	2	3	9	15	10	5	3	2	1	4	3	2	5	2	8	2	7	4	2	3	6	4	3	1	5	4	5	8	6	9	0	1	3	2	3	5	4	6				
TYSON LIVING LEARNING CENTER	4	3	11	1	1	1	8	10	5	5	2	3	3	3	1	2	3	2	8	1	2	10	4	2	5	0	3	1	5	4	5	6	7	8	0	1	3	2	3	5	4	6				
ECO SENSE	4	3	11	1	1	1	9	10	5	5	2	3	3	3	1	2	3	2	4	1	2	10	4	2	5	0	3	1	5	4	5	6	7	8	0	1	3	2	3	5	4	6				
PAINTERS HALL	4	3	11	2	1	1	9	10	5	5	2	3	3	3	2	4	1	9	1	2	9	3	2	3	5	0	3	1	5	4	5	6	7	8	0	1	3	2	3	5	4	6				
IDEAs Z2 DESIGN FACILITY	4	3	11	2	1	1	9	10	5	5	2	3	3	3	2	4	1	9	1	2	9	3	2	3	5	0	3	1	5	4	5	6	7	8	0	1	3	2	3	5	4	6				
<b>PASSIVE HOUSE</b>	4	3	13	1	2	2	7	11	6	5	2	0	2	4	2	2	2	5	3	8	2	2	7	4	3	3	4	1	5	4	5	7	6	8	0	1	3	2	3	5	4	6				
FAMILY HOUSE, HUDSON, US	4	3	8	2	0	2	7	8	4	4	2	0	2	3	3	2	4	3	7	2	2	7	4	3	3	3	3	6	1	3	2	3	3	4	5	6	7	8	0	1	3	2	3	5	4	6
RURAL REGENERATION CENTER, U	2	3	8	0	1	1	6	7	3	4	2	0	2	3	2	1	2	4	3	6	2	2	6	4	3	2	3	0	1	3	2	3	3	4	5	6	7	8	0	1	3	2	3	5	4	6
STANDINGS COURT RESIDENCES, U	3	3	9	1	1	1	7	8	4	4	2	0	2	3	3	1	3	6	4	7	2	2	8	4	4	3	2	6	1	3	2	3	3	4	5	6	7	8	0	1	3	2	3	5	4	6
DAS BIOHAUS, US	3	3	9	1	1	1	7	8	4	4	2	0	2	3	3	1	3	6	4	7	2	2	8	4	4	3	2	6	1	3	2	3	3	4	5	6	7	8	0	1	3	2	3	5	4	6

S:Sources; NoCp: Number of Components; R:renewable; n-R:non renewable; BB:Building Boundary; SB:Site Boundary; Pr:Producer; St:Storage; Cm:Consumer; R-P:Recycle Pathways; C-P:Control Pathways

E1:Envelope Unit; E2:Lighting System Unit; E3:Plugload Equipment Unit; E4:Hydraulic System Unit; E5:Heating System Unit; E6:Hot Water Unit; E7:Users Unit

The analysis of the data showed interesting correlations between the architectural parameters and the energy systems diagrams developed for the buildings in question. Figure 5 presents the relationship between the energy quality ratio and the amount of energy consumed; that is, those buildings that have a higher number of energy transformations occurring within the system tend to use less energy.

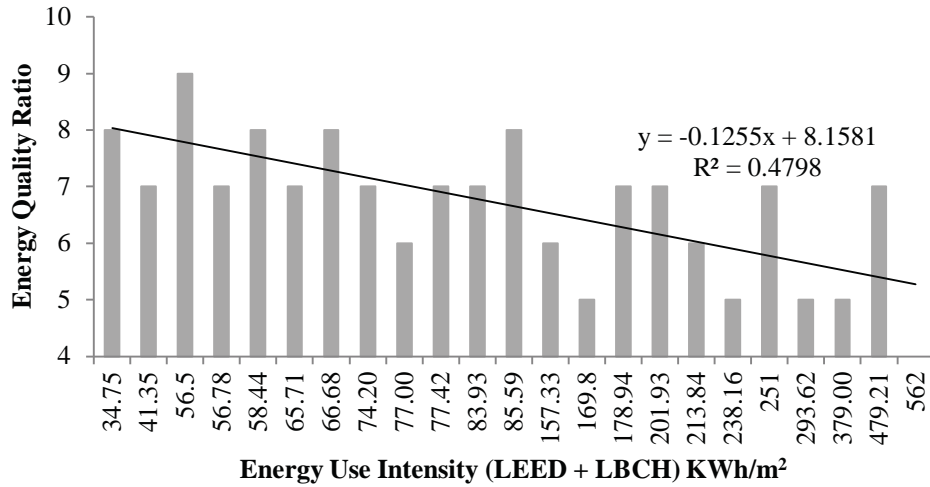


Figure 5: Energy Quality Ratio vs. Energy Use Intensity (EUI). The energy quality ratio is proportional to the EUI with a slope of -0.1.

In Figure 6 there is a stronger trend showing that the energy quality ratio is related to the percentage of renewable energy produced on site; that is, buildings with higher production of renewable energy tend to have a higher number of transformations of energy.



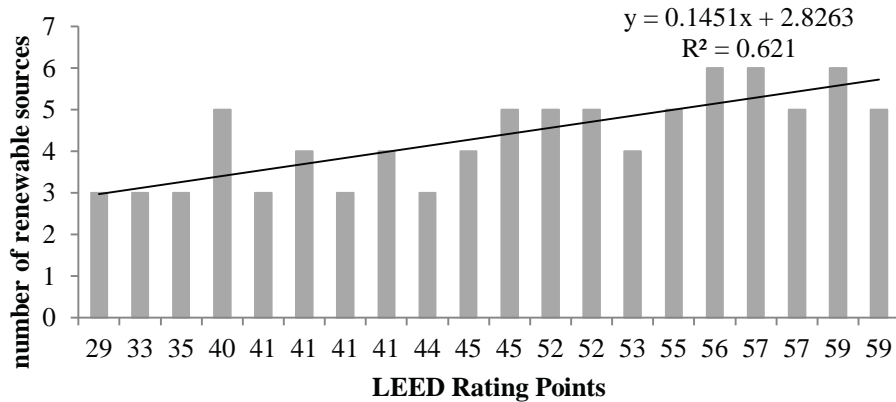


Figure 8: Renewable sources vs. LEED rating points. The number of renewable sources is proportional to the LEED rating points with a slope of 0.1.

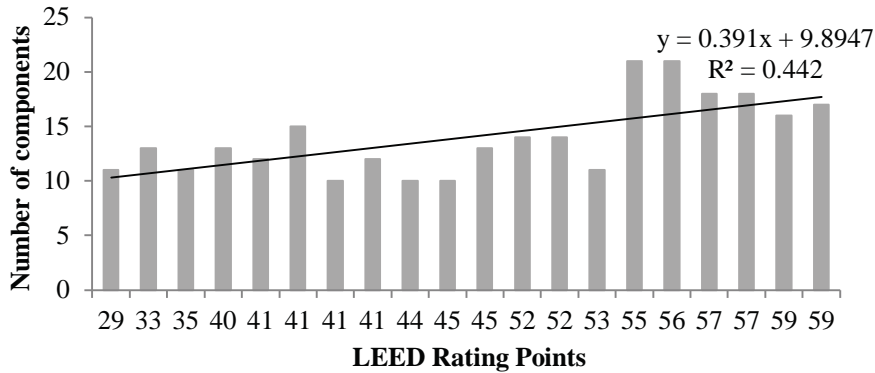


Figure 9: Number of components vs. LEED rating points. The number of components of the systems diagrams is proportional to the LEED rating points with a slope of 0.4.

## Discussion and Conclusion

The drawing of the energy systems diagrams and analysis of the available data indicated that there were differences among the rating systems of these organizations that could affect the graphs shown above. Because the Passive House certification focuses on energy use reduction, the data available on their website mostly related to those strategies that accomplished the rating system requirements. Therefore, little or no data was offered regarding other design strategies. Conversely, the LEED and Living Building Challenge rating systems provided a more holistic approach, and a wider variety of data was available to build the energy system diagrams. However, buildings from the Passive House database showed the lowest levels of energy use when compared to those selected from the other two rating systems. This suggests that the rating systems organizations could benefit from considering the requirements of their peer rating systems. Buildings, like ecosystems, are complex systems, and they could be designed not from a single, reductionist perspective but from a holistic one.

Regardless of the differences among rating systems, several conclusions emerge from this study. The energy systems diagrams graphically represent the flows of energy and the organization of environmental systems; and by doing this they are also representing their underlying patterns (i.e., reiterated behavior). These patterns are characterized using parameters such as energy sources, energy quality, energy flows, energy storage, system structure, and system interactions. In this document, it is shown that these parameters can also be applied to characterize buildings. So the first conclusion is that buildings can be represented using the energy system diagrams.

Moreover, the abovementioned parameters show a correlation with those parameters commonly used to define the environmental performance of buildings; that is, energy use intensity, energy production on-site, and certification levels. This correlation allows us to conclude that the energy system diagrams not only represent buildings, but they also can represent the environmental performance of buildings. For example, a diagram with long energy chains or a diagram with several producer units, are both more likely to correspond to a building that has environmental strategies integrated in its design. However, it is the whole system rather than a single flow, which can give more insight about the environmental performance of a building.

Based on the above statements, the final conclusion is that the energy systems diagrams have potential as a useful platform for characterizing patterns of structure and dynamics that might be used in architectural design. In other words, the energy system diagrams might be employed as a qualitative ecomimetic tool for analyzing energy patterns in ecosystems and for finding correlations with energy patterns in buildings. This analysis, when done with the proposed tool (i.e., the energy system diagrams), allows communication at a very abstract and qualitative level between architecture and ecology. The energy system diagrams can show, for example, how a desert ecosystem manages the flows of water; and by comparing the ecosystem flows and organization with a specific building we can gain more insight about potential ecomimetic strategies. Not all energy ecosystem patterns are likely to be transferred, nor is that the intention of an ecomimetic approach. By using the energy system diagrams a trans-disciplinary language is proposed to advance the construction of an ecomimetic methodology.

Future work will have to be done to illustrate pattern transferring. Also, future work will be focused on developing a quantitative ecomimetic tool in order to measure those parameters in ecosystems patterns to be transferred.

## **Acknowledgements**

This research work is founded by Fundacion Caja Madrid, Spain.



## REFERENCES

- Benyus, J. M. 1997. *Biomimicry : innovation inspired by nature*. New York: Morrow.
- BioTRIZ. 2012. [cited November, 20 2012]. Available from <http://www.biotriz.com/our-methods>.
- Brown, M. T. 2004. "A picture is worth a thousand words: energy systems language and simulation." *Ecological Modelling* no. 178 (1–2):83-100. doi: 10.1016/j.ecolmodel.2003.12.008.
- DTI, (Department of Trade and Industry). 2007. "Biomimetics: strategies for product design inspired by nature – a mission to the Netherlands and Germany. Report of a DTI GlobalWatchMission."
- Gruber, P. 2011. *Biomimetics in architecture architecture of life and buildings*. Vienna; New York: Springer.
- Helms, M., Vattam S. S., and Goel A. K. 2009. "Biologically inspired design: process and products." *Design Studies* no. 30 (5):606-622. doi: 10.1016/j.destud.2009.04.003.
- International Living Future Institute. 2012. *Living Building Challenge Case Studies* [cited August 2012]. Available from <https://ilbi.org/lbc/casestudies/?searchterm=case%20studies>.
- Mitsch, W. J., and S. E. Jørgensen. 2003. "Ecological engineering: A field whose time has come." *Ecological Engineering* no. 20 (5):363-377. doi: 10.1016/j.ecoleng.2003.05.001.
- Odum, H.T. 1994. *Ecological and general systems: an introduction to systems ecology*: University Press of Colorado.
- Odum, H. T. 2007. *Environment, power, and society for the twenty-first century : the hierarchy of energy*. New York: Columbia University Press.
- Odum, H. T., and Peterson, N. 1996. "Simulation and evaluation with energy systems blocks." *Ecological Modelling* no. 93 (1–3):155-173. doi: 10.1016/0304-3800(95)00221-9.
- Passive House Institutions. 2012. Passive House Database [cited August 2012]. Available from <http://www.passivhausprojekte.de/projekte.php>.
- Pedersen, Z. 2007. Biomimetic approaches to architectural design for increased sustainability (paper no. 033).
- Srinivasan, R. S., Braham W. W., Campbell, D. E., and Curcija C. D. Curcija. 2012. "Re(De)fining Net Zero Energy: Renewable Energy Balance in environmental building design." *Building and Environment* no. 47 (0):300-315. doi: 10.1016/j.buildenv.2011.07.010.
- USGBC. 2012. LEED Projects & Case Studies Directory [cited August 2012]. Available from <http://www.usgbc.org/LEED/Project/CertifiedProjectList.aspx>.
- von Bertalanffy, L. 2008. "An Outline of General System Theory." *Emergence: Complexity & Organization* no. 10 (2):103-123.

## ABOUT THE AUTHORS

**Mercedes Garcia-Holguera:** Mercedes Garcia-Holguera was born in Salamanca, Spain and graduated from the Polytechnic University of Madrid, where she studied Architecture. After working for several years in Spain, Chile and Mexico, and receiving a Certificate on Sustainable Architecture and the accreditation as LEED AP, She started her PhD program in 2011 in Bioresource Engineering at McGill University, Canada. Her research work focuses on biomimetics, ecological engineering and regenerative architecture. She is under the supervision of Professor Grant Clark (Bioresource Engineering), Professor Susan Gaskin (Civil Engineering) and Professor Aaron Sprecher (School of Architecture). Mercedes is assessing ecosystem patterns and developing conceptual and quantitative architectural models using ecological engineering tools. Her research will provide the scientific basis for biomimetic and regenerative architecture design.

**Grant Clark:** McGill University, Canada

**Susan Gaskin:** McGill University, Canada

**Prof. Aaron Sprecher:** Aaron Sprecher has been an Assistant Professor in the School of Architecture in the Faculty of Engineering at McGill University since 2008. He is co-founder and partner of Open Source Architecture, an international collaborative research group that brings together leading international researchers in the fields of design, engineering, media research, history and theory. He completed his graduate studies at Bezalel Academy of Art and Design (Israel) and the University of California Los Angeles (UCLA). His research and design work focuses on the synergy between information technologies, computational languages and automated digital systems, examining the way in which technology informs and generates innovative approaches to design processes. Beside numerous publications and exhibitions, he has lectured in many institutions including the Massachusetts Institute of Technology, The Cooper Union, and Harvard University. Aaron Sprecher is co-curator and co-editor of the exhibition and publication “The Gen(H)ome Project” (MAK Center, Los Angeles, 2006) and design curator of “Performatism” (Tel Aviv Museum of Art, 2008). He is a recipient of research grants, most recently, awarded with the prestigious Canada Foundation for Innovation award and Social Sciences and Humanities Research Council (SSHRC) grants.



## Trace Metal Contamination Influenced by Land Use, Soil Age, and Organic Matter in Montreal Tree Pit Soil

Maryam Kargar,\* Pierre Jutras, O. Grant Clark, William H. Hendershot, and Shiv O. Prasher

The short life span of many street trees in the Montreal downtown area may be due in part to higher than standard concentrations of trace metals in the tree pit soils. The effects of land use, soil organic matter, and time since tree planting in a given tree pit (soil age) were studied with respect to the total concentration of trace metals (Cr, Ni, Cu, Zn, Cd, and Pb) in soil collected from tree pits on commercial and residential streets. Contingency table analysis and multiple linear regression were applied to study how these variables were related to the total concentrations of trace metals in soil. Other variables, such as pH, street width, distance of the tree pit from the curb, and tree pit volume, were also used as input to statistical analysis to increase the analysis' explanatory power. Significantly higher concentrations of Cu, Cd, Zn, and Pb were observed in soils from commercial streets, possibly as a result of heavier traffic as compared with residential streets. Soil organic matter was positively correlated with the concentrations of Cu and Pb, probably due to the ability of organic matter to retain these trace metals. Nickel, Cu, Zn, Cd, and Pb were positively correlated with the soil age presumably because trace metals accumulate in the tree pit soil over time. This knowledge can be helpful in providing soil quality standards aimed at improving the longevity of downtown street trees.

**I**N URBAN ENVIRONMENTS, there are often trees planted in sidewalk tree pits along downtown streets. The environmental and social benefits provided by the large number of public trees in the City of Montreal are valued at around \$648 million (Jutras, 2008). However, these trees face conditions that threaten their health and shorten their life expectancy as compared with forest trees (Quigley, 2004). Little is known about the effects of urbanization on soil properties and processes in reference to tree health (Hagan et al., 2012; Wei and Yang, 2010). Hence, identifying the limiting factors for tree health in urban environments can be crucial for urban tree management. Compacted soil and exclusion of air and water from the soil pore space are common problems affecting tree health in sidewalks (Harris et al., 1999a), as are trace metal pollution (Chen et al., 2010; Li et al., 2013) and contamination with deicing salt (Bäckström et al., 2004; Cunningham et al., 2008; Czerniawska-Kusza et al., 2004). The presence of these contaminants in tree pit soil can affect the availability of nutrients for the tree roots (Roberts et al., 2006). This can happen because of changes in the cation exchange capacity of the soil or alterations to the population of mycorrhizal species that help make the nutrients available to the tree roots (Day et al., 2010).

Potential health risks and the disturbance of geochemical cycling in urban ecosystems are other important issues on which urban soil contaminants, such as trace metals, have a direct influence (Xia et al., 2011). Much attention has therefore been given during recent decades to issues related to urban soil contamination. In addition, the soil in tree pits can play a significant role in filtering contaminants from urban runoff water and in improving urban runoff water quality, which is one of the broadly stated goals of stormwater management (U.S. National Research Council Committee on Reducing Stormwater Discharge Contributions to Water Pollution, 2009).

Land use, as a measure of anthropogenic influence, is a governing factor that influences the chemical properties of tree pit soil. Its effect on tree pit soil can be direct (e.g., physical disturbance) or indirect (e.g., soil hydrophobicity as a result of deicing salt application) (Pouyat et al., 2007). According to the type of urban land use (e.g., commercial vs. residential), the direct or indirect effects of land use on physical and chemical

Copyright © American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America. 5585 Guilford Rd., Madison, WI 53711 USA. All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher.

J. Environ. Qual. 42:1527–1533 (2013)

doi:10.2134/jeq2013.02.0055

Received 17 Feb. 2013.

\*Corresponding author (maryam.kargar@mail.mcgill.ca).

M. Kargar, P. Jutras, O.G. Clark, and S.O. Prasher, Dep. of Bioresource Engineering, Macdonald Campus, McGill Univ., 21111 Lakeshore Rd., Ste-Anne-de-Bellevue, QC, Canada H9X 3V9; W.H. Hendershot, Dep. of Natural Resource Sciences, Macdonald Campus, McGill Univ., 21111 Lakeshore Road, Ste-Anne-de-Bellevue, QC, Canada H9X 3V9.

**Abbreviations:** SOM, soil organic matter.

properties of urban soils will be different (Harris et al., 1999b; Park et al., 2010).

In terms of the effect of land use on trace metal concentration in urban soil, different studies have indicated different results. Pouyat et al. (2007) studied the effects of land use on trace metal concentration in soil in Baltimore, Maryland. The land use categories considered were commercial, industrial, institutional, transportation rights-of-way, high- and medium-density residential, golf course, park, urban open, unmanaged forest, and wetland. The results indicated that trace metal concentrations (Cu, Cr, Ni, Pb, and Zn) were not related to land use. The lack of statistically significant differences in that study may have been due to the scale of observation and the sampling intensity. In contrast, Guo et al. (2012) observed significantly higher concentrations of trace metals (Pb, Zn, and Cu) in commercial zones as compared with residential zones in Yibin City, Sichuan Province, China. In commercial zones, the vehicular traffic led to increased concentrations of trace metals, such as Pb, Cd, and Cu, in the soil along the street (Czarnowska, 1999) because the metals are carried in particles from vehicle exhaust, the wear of brakes and tires, etc. (Wei and Yang, 2010).

The fate of the trace metals in the soil depends on many soil processes, which are mainly controlled by pH and the type and amount of soil organic matter (Kabata-Pendias, 2001a). The soil pH is an important parameter affecting trace metal sorption-desorption, precipitation-dissolution, complex formation, and oxidation-reduction reactions (McLean and Bledsoe, 1992). Organic matter also has an important effect on trace metal transportation and accumulation in soil contaminated with trace metals (Kabata-Pendias, 2001b). However, tree pit soil does not generally have rich top soil (Jim, 1998). Limited organic matter additions to the tree pit soil over time (Roberts et al., 2006) and organic matter dispersion under the effect of deicing salt in the urban environment (Green et al., 2008) may cause a decrease in the organic matter content of the soil.

The heterogeneity of urban soil chemistry cannot be attributed only to the differences in land use. Researchers who previously examined soils under similar land use regimes found that key soil chemical properties can also have significant relationship with the soil age (Park et al., 2010). In the current study, the age of the tree pit soil is considered as the time since the trees were transplanted from the nursery to the tree pit. At the end of the lifespan of a street tree, the dead tree is removed, and the previous soil in the tree pit is excavated. Fresh soil is added when the new young tree is planted in the tree pit. Over time, trace metal concentrations in the soil can increase due to traffic emissions and runoff water (Peltola et al., 2005).

The main objective of this study is the chemical characterization of soil in Montreal tree pits exposed to urban runoff for 4 to 40 yr. The results of this study can improve our understanding of the impact of different environmental and soil chemical characteristics on trace metal concentrations in tree pit soil. It is likely that most types of urban soil, including tree pit soil, contain higher than background levels of trace metals (Craul, 1992). Therefore, the intent in this study was to answer the following questions: (i) What is the trace metal (Cr, Ni, Cu, Zn, Cd, and Pb) concentration present in the soil of sidewalk tree pits in Montreal? (ii) Do these trace metal concentrations exceed standards recommended by the Canadian Council of Ministers of the Environment (1999) for urban soil in commercial and residential zones, and, if so, what

factors are correlated with these excess concentrations? (iii) Do land use, soil organic matter (SOM), and soil age have significant relationships with trace metal (Cr, Ni, Cu, Zn, Cd, and Pb) concentrations in tree pit soils?

## Materials and Methods

### Study Site Description

The study was conducted in the City of Montreal (Quebec, Canada) (45°30' N, 73°34' W), which has a humid continental climate characterized by hot summers and cold winters. The mean annual precipitation and snowfall from 1971 to 2000 were 106 and 226 cm, respectively (Environment Canada, 2012a). The average summer and winter temperatures from 1981 to 2010 were around 19 and  $-8^{\circ}\text{C}$ , respectively (Environment Canada, 2012b). The study area was classified into two land use zones: commercial and residential. Commercial zones include heavily used streets, and tree growth rates are only slow to fair. Residential zones are used only by residents, and trees are typically vigorous (Jutras et al., 2010). In addition to the importance of vehicular and pedestrian traffic and tree growth rate, other criteria for land use classification included the size of the tree pits, the level of exposure to deicing salts, and the width of the street (Jutras et al., 2010). Five streets in this study were categorized as lying in commercial zones, and 11 streets were in residential zones.

### Soil Sampling

During a prior study, Jutras (2008) collected four soil samples at the edges of each of 796 tree pits (at a depth of 0–20 cm) on the selected streets in five land use zones (intensive commercial, commercial, institutional, intensive residential, and residential) in the City of Montreal and mixed them to form a composite for each tree pit. The samples were dried at room temperature for 14 d, sieved through a 2-mm metal sieve to remove large particles, and stored in closed plastic bags until analysis. In the current study, land use, SOM, and soil age were considered as the three most important independent variables affecting the chemistry of the tree pit soil. A total of 73 samples were selected randomly from the original sample set, including 35 samples from commercial zones and 38 samples from residential zones.

### Independent Variable Classes for Sample Selection

For the current study, samples were selected only from commercial and residential zones because they were more numerous and were expected to show greater contrast. The samples were sorted according to SOM and soil age classes. At this step, considering the acceptable range (4–7%) of SOM in the standard recipe for Montreal tree pit soil, samples with SOM content higher than 10% were omitted from the sample set, and the rest were classified into two groups: samples with  $\text{SOM} \leq 5.5\%$  and samples with  $5.5 < \text{SOM} \leq 10\%$ . The samples in each SOM class were then clustered into four age classes, partitioned to achieve an adequate number of samples in each class: age < 12 yr,  $12 \leq \text{age} < 17$  yr,  $17 \leq \text{age} < 28$  yr, and  $28 \text{ yr} \leq \text{age}$ . Every class contained three to five samples, for a total of 73 samples.

### Analysis

#### Chemical Analysis

After sample selection, different chemical characteristics of the soil samples were measured. Soil pH was measured by a pH meter



(Accumet AR10; Fisher Scientific Inc.) in water using a 1:2 (7 g soil: 14 mL H<sub>2</sub>O) soil-to-solution ratio (Hendershot et al., 1993). Total concentrations of trace metals, including Cr, Ni, Cu, Zn, Cd, and Pb, were estimated by hot nitric acid digestion, wherein 2 mL of concentrated nitric acid (16 mol L<sup>-1</sup>) was added to a 0.2-g soil sample, heated for 5 h on a block heater set at 130°C, cooled, diluted up to 50 mL with nanopure water, and left to settle overnight. The total concentrations of trace metals were then quantified in the extracts using an inductively coupled plasma–mass spectrometer (Varian 820 MS) (Hendershot et al., 2008). Quality control of the analysis of total concentration of trace metals was assured by analysis of certified reference material (SED 90–4 and SED 94–3 provided by the program Environment Canada Proficiency Testing). Recovery percentages of the SED 90–4 and SED 94–3 were as follows: 99.08 and 105.99% for Cr; 108.83 and 98.17% for Ni; 106.82 and 100.36% for Cu; 101.12 and 101.67% for Zn; 97.46 and 99.05% for Cd; and 105.50 and 101.20% for Pb.

### Statistical Analysis

Contingency table analysis and multiple linear regression were done using SPSS Statistics (V.20, IBM). The data were normalized using a base-10 logarithmic transformation, and normality was confirmed with the Kolmogorov-Smirnov test. Multiple linear regression was used to determine the relationships between dependent variables (trace metal concentrations: Cr, Ni, Cu, Zn, Cd, and Pb) and independent variables (land use, SOM, and soil age). It was also applied to explore whether the inclusion of other variables (pH, street width, distance of the tree pit from the curb, and tree pit volume [measured in Jutras, 2008]) could increase the explanatory power of the statistical model.

A backward elimination method was used to determine which independent variables were important. This method starts with all candidate independent variables and then eliminates the variables that contribute the least to explain the variation of the dependent variable. This process is repeated until all the remaining independent variables in the model have a significant partial regression coefficient (Legendre and Legendre, 1998).

To exclude collinearity between the regression factors, the variance inflation factor was computed; generally a value greater than 10 indicates collinearity between variables (O'Brien, 2007).

**Table 1. Classes of trace metal concentrations for contingency table analysis.**

Trace metals	Upper class boundary†			
	Class 1	Class 2	Class 3	Class 4
	mg kg <sup>-1</sup> soil			
Cr	33	54	66	100
Ni	16	22	34	100
Cu	53	80	216	
Zn	123	255	950	
Cd	0.4	0.8	6.2	
Pb	100	290	670	2151

† Each class includes the range of trace metal concentrations smaller or equal to the upper class boundary (the first class range starts at 0).

The *R*<sup>2</sup> value was calculated to indicate the overall predictive accuracy of the multiple linear regression model.

Analysis of variance using the general linear model was applied to characterize the interactions between land use classes and other independent variables. Hierarchical cluster analysis (Ward's minimum variance method) was used to categorize the dependent variables for contingency table analysis (Table 1), which was then used to show the effect of land use (a categorical variable) on total concentration of trace metals. Ward's minimum variance method is used to group data by minimizing the total intercluster variance while maximizing the intracluster variance (Brown and Martin, 1996).

## Results and Discussion

### Descriptive Statistics

As descriptive statistics of the soil chemical parameters indicated (Table 2), Cd and Pb in commercial zones and Pb in residential zones demonstrated coefficients of variation higher than 1. The soil Zn and Cu also indicated high coefficients of variation values in the commercial zone. This can be attributed to the nonhomogeneous distribution of metals in tree pit soils (Li et al., 2013) or to their anthropogenic source, such as vehicle emissions, oxidation of car lubricants, and gradual deterioration of vehicle mechanical parts (Chen et al., 2012).

The results of the Kolmogorov-Smirnov test (*P* ≤ 0.05) of data from commercial and residential zones showed that Cd

**Table 2. Descriptive statistics of the soil chemical properties in commercial and residential zones.**

Variable	Commercial zone (n = 35)						Residential zone (n = 38)					
	Min.	Max.	Avg.	SD	CV†	Standard‡	Min.	Max.	Avg.	SD	CV†	Standard§
pH	6.73	8.41	7.55	0.39	0.05	6–8	6.52	8.10	7.44	0.35	0.05	6–8
SOM,¶ %	2.16	9.71	5.91	2.06	0.35	4–7¶	3.31	9.55	5.98	1.80	0.30	4–7#
Cr, mg kg <sup>-1</sup>	14.5	98.6	42.0	24.2	0.58	87	16.2	66.3	31.7	12.4	0.39	64
Ni, mg kg <sup>-1</sup>	8.94	97.6	24.6	17.6	0.72	50	9.44	37.3	20.2	6.25	0.31	50
Cu, mg kg <sup>-1</sup>	9.83	216	72.7	57.8	0.79	91	9.93	81.6	31.2	15.4	0.50	63
Zn, mg kg <sup>-1</sup>	33.9	948	262	222	0.85	360	51.7	395	127	71.6	0.56	200
Cd, mg kg <sup>-1</sup>	0.13	6.13	1.08	1.32	1.22	22	0.19	1.28	0.48	0.27	0.56	10
Pb, mg kg <sup>-1</sup>	13.7	2150	357	516	1.44	260	15.7	678	93.2	119	1.28	140

† Coefficient of variation.

‡ Canadian soil quality guidelines for protection of the environment and human health in a commercial zone. Source: Canadian council of Ministries of the Environment (1999).

§ Canadian soil quality guidelines for protection of the environment and human health in a residential zone. Source: Canadian council of Ministries of the Environment (1999).

¶ Soil organic matter.

# City of Montreal's recipe for the soil mixtures used in tree pits. Source: City of Montreal (1995).

and Pb concentrations were not normally distributed. Hence, all the data were normalized with a base-10 logarithmic transformation.

Figure 1 indicates the sample frequency percentage across chemical parameters, which fall below, are equal to, or exceed the standard values in commercial and residential zones. These standard values were derived from Canadian soil quality guidelines for protection of the environment and human health in commercial and residential zones (Canadian Council of Ministers of the Environment, 1999). The soil Cd concentration in commercial and residential zones and the soil Ni concentration in the residential zone did not exceed the standard values. In commercial zones, Cr and Ni concentrations were not significantly different from the standard values, whereas pH, SOM, Cu, Zn, and Pb concentrations were significantly higher than the standard values.

As Wei and Yang (2010) reported, in general trace metals such as Cu, Pb, and Zn show the highest contamination in urban soil in comparison with other trace metals, such as Cr and Ni. In the residential zone, pH, Cr, and Cu were not significantly different from the standards. Among different soil chemical parameters, only SOM fell significantly below the recommended value for Montreal tree pit soils in commercial ( $p \leq 0.01$ ) and residential zones ( $p \leq 0.05$ ). According to Jim (1998), the inadequacy of organic matter is a common problem for urban soils and may be attributed to different factors, such as the low amount of litter returned to the soil (Roberts et al., 2006) and organic matter dispersion (Green et al., 2008) and leaching from the soils exposed to deicing salt (Green and Cresser, 2008).

## The Effect of Land Use

According to the measured chemical parameters of the soil samples (Table 3) and the result of contingency table analysis (Table 4), Cu, Zn, Cd, and Pb concentrations were significantly higher in commercial zones than in residential zones, as corroborated by other studies. Considering the traffic emissions as one of the main possible sources of these trace metals in urban soil (Wei and Yang, 2010), this significantly higher concentration of trace metals in the commercial zones can be attributed to the heavier traffic in the commercial streets in comparison to the residential ones (Mingkui and Hao, 2009).

## The Effects of Other Independent Variables

Analysis of variance for Zn did not show any significant interactions between land use and any of the other independent variables (organic matter, pH, age, street width, distance of the tree pit to the curb, and tree pit volume) (Table 5). Concentrations of Cr and Ni, however, were significantly influenced by land use  $\times$  street width, land use  $\times$  distance of the tree pit to the curb, and land use  $\times$  tree pit volume. The concentration of Cu was significantly influenced by land use  $\times$  SOM and land use  $\times$  street width. The concentration of Cd was influenced by land use  $\times$  SOM, and that of Pb was influenced by land use  $\times$  SOM and

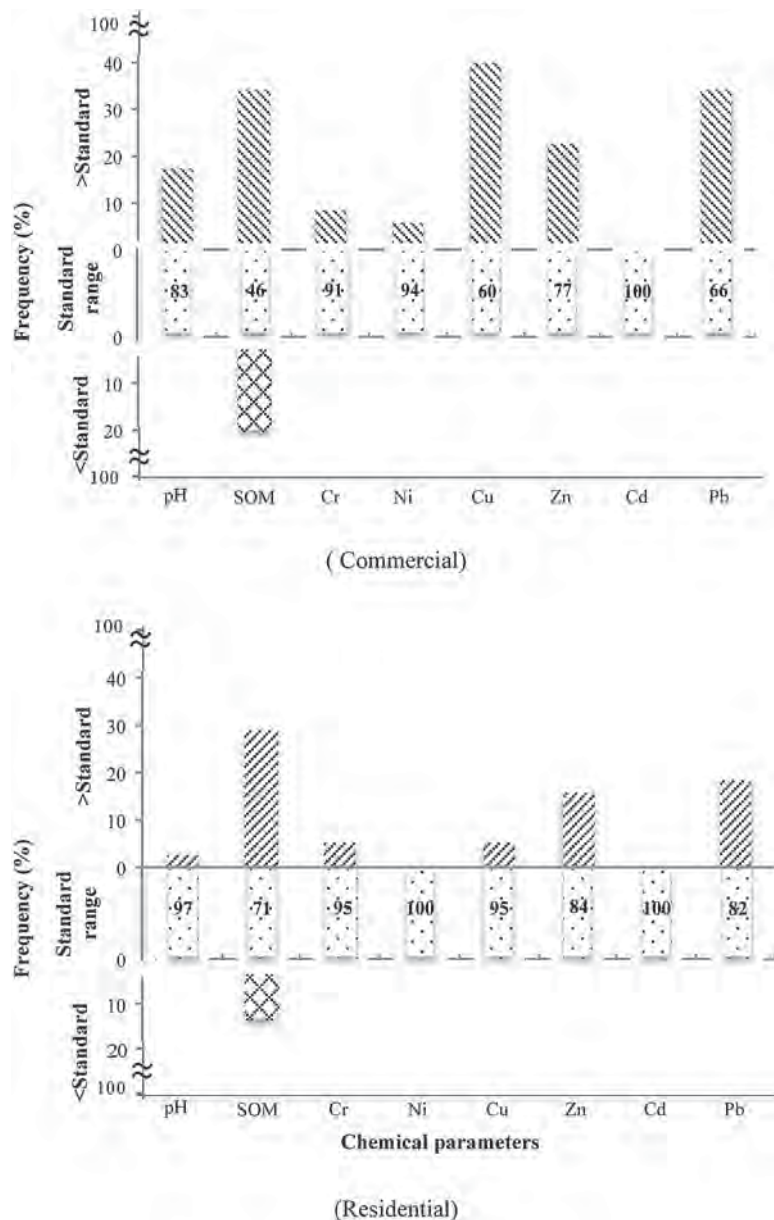


Fig. 1. Frequency (%) of soil samples that were below, within, or above the recommended range of values for a number of chemical parameters in commercial and residential zones. The recommended range of values (standards) refer to the Canadian soil quality guidelines for protection of the environment and human health in commercial and residential zones. The symbol  $\approx$  shows the y axis scale break to illustrate the small frequencies more clearly.

land use  $\times$  tree pit volume. Therefore, samples from commercial and residential zones were examined separately.

Multiple linear regression was applied to evaluate how the other independent variables played a role in determining how the values of the dependent variables changed in the commercial and residential zones (Table 6). All the variance inflation factor values were below 10, which negates the possibility of collinearity among the independent variables in every model.

It seems that variations in environmental factors, such as tree pit volume or distance to the curb, can result in spatial heterogeneity of the tree pit soil and might obscure expected relationships between the other variables of interest in the tree pits. However, there are general identifiable patterns of the effect that urban environments have on concentrations of trace metals in the soil.

**Table 3. Soil chemical parameters as related to independent variables.**

Land use	SOM†	Age	n	pH	Cr	Ni	Cu	Zn	Cd	Pb
	%									
Commercial	≤5.5	x < 12	3	7.47 (0.34)‡	31.0 (7.2)	17.5 (1.8)	36.9 (20.8)	96.5 (27.9)	0.31 (0.09)	40.2 (6.4)
		12 ≤ x < 17	5	7.56 (0.19)	26.4 (5.2)	15.9 (1.8)	36.5 (11.9)	155 (48)	0.49 (0.18)	211 (100)
		17 ≤ x < 28	5	7.81 (0.16)	40.8 (7.2)	23.1 (2.1)	96.3 (7.6)	466 (132)	2.05 (0.88)	592 (309)
		28 ≤ x	3	7.96 (0.14)	75.2 (8.3)	30.0 (0.9)	153 (19)	441 (30)	3.01 (1.56)	1060 (199)
	5.5 < x ≤ 10	x < 12	5	7.42 (0.10)	45.0 (12.0)	18.5 (2.5)	68.8 (23.7)	177 (44)	0.62 (0.14)	106 (21)
		12 ≤ x < 17	5	7.19 (0.14)	22.5 (4.2)	11.5 (1.1)	15.6 (3.0)	72.1 (15.3)	0.26 (0.07)	29.2 (5.9)
		17 ≤ x < 28	5	7.58 (0.11)	34.8 (7.9)	23.4 (3.9)	47.0 (14.3)	312 (151)	0.92 (0.36)	147 (74)
		28 ≤ x	4	7.54 (0.26)	75.9 (11.1)	63.9 (13.1)	163 (18)	409 (36)	1.55 (0.08)	945 (403)
Residential	≤5.5	x < 12	3	7.60 (0.26)	20.8 (2.5)	18.3 (5.6)	18.7 (4.5)	71.0 (11.8)	0.25 (0.05)	29.9 (7.5)
		12 ≤ x < 17	5	7.57 (0.09)	28.8 (4.0)	17.8 (1.6)	34.1 (9.7)	165 (65)	0.60 (0.15)	104 (36)
		17 ≤ x < 28	5	7.42 (0.23)	36.0 (3.3)	24.8 (1.9)	33.9 (3.6)	121 (14)	0.45 (0.06)	58.0 (7.3)
		28 ≤ x	5	7.61 (0.01)	27.2 (3.3)	18.6 (2.0)	17.8 (1.1)	80.2 (5.3)	0.30 (0.01)	36.4 (3.0)
	5.5 < x ≤ 10	x < 12	5	7.13 (0.13)	23.0 (3.1)	14.2 (1.4)	26.6 (3.4)	83.5 (6.9)	0.31 (0.02)	26.8 (4.8)
		12 ≤ x < 17	5	7.40 (0.08)	41.4 (7.7)	24.2 (3.6)	35.2 (6.4)	146 (29)	0.64 (0.18)	209 (121)
		17 ≤ x < 28	5	7.18 (0.15)	40.3 (8.1)	23.9 (3.4)	41.3 (5.6)	166 (18)	0.60 (0.09)	118 (21)
		28 ≤ x	5	7.69 (0.16)	31.8 (4.2)	19.2 (1.2)	36.5 (11.3)	159 (34)	0.58 (0.15)	138 (54)

† Soil organic matter.

‡ Values are the mean (SE) of the number of samples (n) shown.

**Table 4. Significance of relationship between trace metal concentration and land use category in contingency table analysis.**

Trace metals	No. of classes†	Pearson $\chi^2$
Cr	4	7.26
Ni	4	6.03
Cu	3	16.1**
Zn	3	17.4**
Cd	3	9.18*
Pb	4	10.8*

\* Significant at the 0.05 probability level.

\*\* Significant at the 0.01 probability level.

† Number of classes of trace metal concentrations.

Organic matter showed a positive relationship with concentrations of Cu and Pb in the residential zone (Table 6). This positive relationship may be due to the high affinity of these trace metals for organic ligands in the soil (Kabata-Pendias, 2001a; McLean and Bledsoe, 1992), agreeing with work elsewhere (e.g., Bäckström et al., 2004). In general, organic matter plays a crucial role in the relationship with trace metals because it tends to form soluble or insoluble complexes with trace metals. Therefore, these complexes can be mobile in the soil profile or retained in the soil (Vega et al., 2004).

Organic matter did not show a significant effect on trace metal concentrations in the commercial zone. Kumar and Edward (2009) related this to the adsorption of trace metals to other adsorbents in the soil, such as Fe and Mn oxyhydroxides and the formation of Cu and Pb precipitates (McBride, 1994). In addition, the insignificance of the relationship between organic matter and trace metals may be ascribed to other factors. These factors include competition among trace metals and other cations for complexing sites on SOM (Katip et al., 2012; Luoma and Davis, 1983; Peng et al., 2009), such as Na from deicing salt, which is more heavily used in the commercial zone, and Ca from the weathering of concrete in buildings and sidewalks (Cunningham et al., 2008). Another factor is the dispersion of organic matter due to high concentrations of deicing salt (Bäckström et al., 2004; Green et al., 2008). These reasons presumably explain the insignificance of the relationships between SOM and trace metal concentrations in the commercial zone.

There is a positive relationship between trace metals in the soil (Ni, Cu, Zn, Cd, and Pb) and soil age. According to Pezzente (1997), this relationship can be attributed to the continual input of these metals to the tree pit soil over time.

Wider streets in the commercial zone can correspond to higher traffic volumes, which might be associated with higher

**Table 5. The significance of the effect of the interactions between land use and other independent variables on the concentration of trace metals in soil.**

Interaction†	Cr	Ni	Cu	Zn	Cd	Pb
Land use × SOM	0.61‡	0.43	0.03*	0.12	0.05*	0.03*
Land use × pH	0.99	0.68	0.92	0.55	0.91	0.92
Land use × age	0.66	0.22	0.6	0.67	0.44	0.91
Land use × StW	0.00**	0.00**	0.03*	0.23	0.26	0.67
Land use × DisSt	0.00**	0.04*	0.14	0.97	0.69	0.29
Land use × PitV	0.02*	0.02*	0.09	0.3	0.17	0.01**

\* Significant at the 0.05 probability level.

\*\* Significant at the 0.01 probability level.

† DisSt, distance of the tree pit to the curb; PitV, tree pit volume; StW, street width.

‡ Values are P values gained from ANOVA using the general linear model.

**Table 6. Predictive models for concentrations of trace metals in commercial and residential zones.**

	Commercial zone (n = 35)†							Residential zone (n = 38)								
	Intercept	Log SOM	Log age	Log pH	Log StW	Log DisSt	Log PitV	R <sup>2</sup>	Intercept	Log SOM	Log Age	Log pH	Log StW	Log DisSt	Log PitV	R <sup>2</sup>
Log Cr					1.56‡	-1.69		0.65***	2.62		0.20		-1.05	0.94	-0.69	0.48***
Log Ni	-0.74		0.36		1.53			0.65***	2.30				-0.77	0.71	-0.46	0.31**
Log Cu	-3.02		0.32		4.07			0.82***	-0.22	0.45			1.48	0.54	-0.35	0.39**
Log Zn	-4.07		0.50	4.27	1.88			0.62***	1.75		0.37				-0.50	0.23**
Log Cd	-4.09		0.55		3.07			0.63***	-0.61		0.31				-0.46	0.20*
Log Pb	-7.84		0.83	9.88			1.16	0.55***	-6.19	1.24	0.72	7.43			-1.09	0.44***

\* Significant at the 0.05 probability level.

\*\* Significant at the 0.01 probability level.

\*\*\* Significant at the 0.001 probability level.

† DisSt, distance of the tree pit to the curb; PitV, tree pit volume; SOM, soil organic matter; StW, street width.

‡ Numbers represent significant partial regression coefficients, which indicate the importance of the relationship between the corresponding dependent and independent variables. Blank cells indicate insignificant relationships.

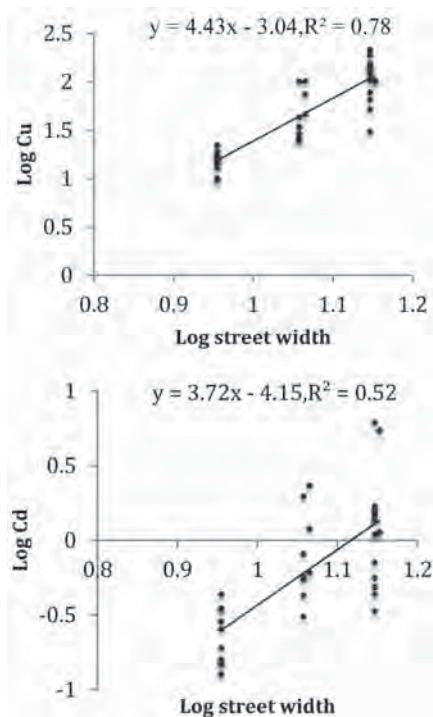
concentrations of trace metals (Jutras et al., 2009). This may explain the significant positive relationship between street width and Cr, Ni, Cu, Zn, and Cd concentrations in the commercial zone (Table 6; Fig. 2). In the residential zones, there is a negative relationship between street width and the concentrations of some trace metals, such as Cr and Ni (Table 6). These relationships may be attributed to the manner in which cars in the residential zone are parked next to the sidewalk, resulting in less splashing, a mechanism by which significant loads of trace metals would otherwise be transported into the tree pit soils (Sabin et al., 2005).

There is also a positive relationship between pH and Pb and Zn concentrations in the soil (Table 6; Fig. 3), as confirmed by Yobouet et al. (2010). In general, maximum retention of cationic metals (e.g., Pb<sup>2+</sup> and Zn<sup>2+</sup>) by the soil particles occurs in pH > 7 (McLean and Bledsoe, 1992). As trace metals are sorbed to the soil solid phase due to an increase in pH, trace metal bioavailability

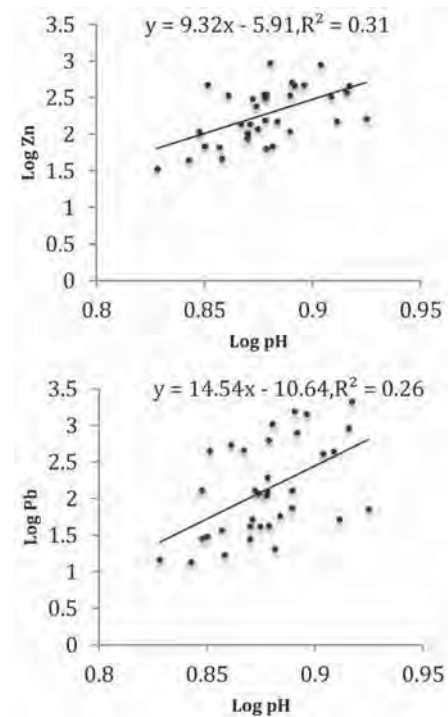
in the soil solution may decrease. In alkaline and neutral environments, Pb and Zn solubility in the soil solution decreases markedly due to chemisorption on oxides and aluminosilicates, complexation with organic matter, and precipitation as carbonates (McBride, 1994; Takáč et al., 2009). Therefore, increasing pH may result in higher Pb and Zn concentrations in the soil solid phase.

### Conclusions

Concentrations of Cu, Zn, and Pb in soils from commercial zones and Zn and Pb concentrations in soils from residential zones were significantly higher than standard values recommended in the Canadian soil quality guidelines for protection of the environment and human health in commercial and residential zones. Heavy vehicular traffic on the wider streets in commercial zones and continual input of trace metals to the tree pit soil over time are some of the mechanisms presumed to result in these



**Fig. 2.** Change in the Log Cu and Log Cd concentrations as related to street width in commercial zones.



**Fig. 3.** Change in the Log Zn and Log Pb concentrations as related to pH in commercial zones.



elevated values. Some soil chemical characteristics, such as SOM and pH, also affected the concentrations of trace metals in the tree pit soil. The results of this study provide a database of the type and rate of the accumulation of trace metals in the tree pit soil. A step forward would be lab and field trials with various soil remediation methods aimed at attenuating excessive trace metal concentrations in the tree pit soil as compared with urban soil quality guidelines. Meeting these guidelines can provide a healthier soil medium for the street trees while increasing soil capacity for runoff infiltration and trace metal retention.

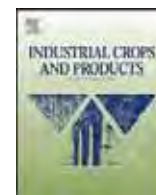
## Acknowledgments

The authors thank the City of Montreal Transport Department for providing the financial support for this project and H el ene Lalande for support during the laboratory work.

## References

- B ackstr om, M., S. Karlsson, L. B ackman, L. Folkesson, and B. Lind. 2004. Mobilisation of heavy metals by deicing salts in a roadside environment. *Water Res.* 38:720–732. doi:10.1016/j.watres.2003.11.006
- Brown, R.D., and Y.C. Martin. 1996. Use of structure-activity data to compare structure-based clustering methods and descriptors for use in compound selection. *J. Chem. Inf. Comput. Sci.* 36(3):572–584. doi:10.1021/ci9501047. doi:10.1021/ci9501047
- Canadian Council of Ministers of the Environment. 1999. Canadian environmental quality guidelines. [http://www.ccme.ca/publications/ccqg\\_rcqg.html](http://www.ccme.ca/publications/ccqg_rcqg.html) (accessed 1 Dec. 2012).
- Chen, X., X. Lu, and G. Yang. 2012. Sources identification of heavy metals in urban topsoil from inside the Xi'an Second Ringroad, NW China using multivariate statistical methods. *Catena* 98:73–78. doi:10.1016/j.catena.2012.06.007
- Chen, X., X. Xia, Y. Zhao, and P. Zhang. 2010. Heavy metal concentrations in roadside soils and correlation with urban traffic in Beijing, China. *J. Hazard. Mater.* 181:640–646. doi:10.1016/j.jhazmat.2010.05.060
- City of Montreal. 1995. Soil standards for transplantation of trees in Montreal tree pits. Standard #2–484. City of Montreal, Montreal, Quebec, Canada.
- Craul, P.J. 1992. *Urban soil in landscape design*. John Wiley & Sons, Hoboken, NJ.
- Cunningham, M.A., E. Snyder, D. Yonkin, M. Ross, and T. Elsen. 2008. Accumulation of deicing salts in soils in an urban environment. *Urban Ecosyst.* 11(1):17–31. doi:10.1007/s11252-007-0031-x
- Czarnowska, K. 1999. Heavy metals in lawn soils of Warsaw. *Soil Science Ann. Poland* 50:31–39.
- Czerniawska-Kusza, I., G. Kusza, and M. Duzyrski. 2004. Effect of deicing salts on urban soils and health status of roadside trees in the Opole Region. *Environ. Toxicol.* 19:296–301. doi:10.1002/tox.20037.
- Day, S.D., P.E. Wiseman, S.B. Dickinson, and J.R. Harris. 2010. Tree root ecology in the urban environment and implications for a sustainable rhizosphere. *Arboric. Urban For.* 36:193–205.
- Environment Canada. 2012a. McGill station mean meteorological statistics, 1971–2000. <http://www.climate.weather-office.ec.gc.ca/> (accessed 15 Dec. 2012).
- Environment Canada. 2012b. Observed climatological average and threshold values (1981–2010) for temperature. <http://weatheroffice.gc.ca/> (accessed 28 Dec. 2012).
- Green, S.M., and M.S. Cresser. 2008. Nitrogen cycle disruption through the application of de-icing salts on upland highways. *Water Air Soil Pollut.* 188:139–153. doi:10.1007/s11270-007-9530-x
- Green, S.M., R. Machin, and M.S. Cresser. 2008. Long-term road salting effects on dispersion of organic matter from roadside soils into drainage water. *Chem. Ecol.* 24(3):221–231. doi:10.1080/02757540802032181. doi:10.1080/02757540802032181
- Guo, G., F. Wu, F. Xie, and R. Zhang. 2012. Spatial distribution and pollution assessment of heavy metals in urban soils from southwest China. *J. Environ. Sci. (China)* 24:410–418. doi:10.1016/S1001-0742(11)60762-6
- Hagan, D., C. Dobbs, N. Timilsina, F. Escobedo, G.S. Toor, and M. Andreu. 2012. Anthropogenic effects on the physical and chemical properties of subtropical coastal urban soils. *Soil Use Manage.* 28:78–88. doi:10.1111/j.1475-2743.2011.00379.x
- Harris, R.W., J.R. Clark, and N.P. Matheny. 1999a. Nutrient management. In: *Arboriculture, integrated management of landscape trees, shrubs, and vines*. 3rd ed. Prentice Hall, Upper Saddle River, NJ, p. 332–371.
- Harris, R.W., J.R. Clark, and N.P. Matheny. 1999b. Planting site: Soil and water. In: *Arboriculture, Integrated Management of Landscape Trees, Shrubs, and Vines*. 3rd ed. Prentice Hall, Upper Saddle River, NJ, p. 95–126.
- Hendershot, W.H., H. Lalonde, and M. Duquette. 1993. Soil reaction and exchangeable acidity. In: M.R. Carter, editor, *Soil sampling and methods of analysis*. Canadian Society of Soil Science, Panawa, Manitoba, Canada, p. 141–143.
- Hendershot, W.H., H. Lalonde, D. Reyes, and J.D. MacDonald. 2008. Trace element assessment. In: M.R. Carter, editor, *Soil sampling and methods of analysis*. 2nd ed. Canadian Society of Soil Science, Panawa, Manitoba, Canada, p. 109–119.
- Jim, C.Y. 1998. Urban soil characteristics and limitations for landscape planting in Hong Kong. *Landsc. Urban Plan.* 40:235–249. doi:10.1016/S0169-2046(97)00117-5
- Jutras, P. 2008. Modeling of urban tree growth with artificial intelligence and multivariate statistics. PhD diss. McGill University, Montreal, Quebec, Canada.
- Jutras, P., S.O. Prasher, and P. Dutilleul. 2009. Identification of significant street tree inventory parameters using multivariate statistical analyses. *Arboric. Urban For.* 35:53–62.
- Jutras, P., S.O. Prasher, and G.R. Mehuys. 2010. Appraisal of key abiotic parameters affecting street tree growth. *Arboric. Urban For.* 36:1–10.
- Kabata-Pendias, A. 2001a. The anthroposphere. In: *Trace elements in soils and plants*. 3rd ed. CRC Press, Boca Raton, FL, p. 3–27.
- Kabata-Pendias, A. 2001b. Soils and soil processes. In: *Trace elements in soils and plants*. 3rd ed. CRC Press, Boca Raton, FL, p. 27–49.
- Katip, A., F. Karaer, S. Ileri, S. Sarmasik, N. Aydogan, and S. Zenginay. 2012. Analysis and assessment of trace elements pollution in sediments of Lake Ulubat, Turkey. *J. Environ. Biol.* 33:961–968.
- Kumar, S.P., and J.K.P. Edward. 2009. Assessment of metal concentration in the sediment cores of Manakudy estuary, south west coast of India. *Indian J. Mar. Sci.* 38:235–248.
- Legendre, P., and L. Legendre. 1998. Interpretation of ecological structures. In: *Numerical ecology*. 2nd ed. Elsevier Science, New York, p. 481–574.
- Li, X., L. Liu, Y. Wang, G. Luo, X. Chen, X. Yang, M.H.P. Hall, R. Guo, H. Wang, J. Cui, and X. He. 2013. Heavy metal contamination of urban soil in an old industrial city (Shenyang) in Northeast China. *Geoderma* 192:50–58. doi:10.1016/j.geoderma.2012.08.011
- Luoma, S.N., and J.A. Davis. 1983. Requirements for modeling trace metal partitioning in oxidized estuarine sediments. *Mar. Chem.* 12:159–181. doi:10.1016/0304-4203(83)90078-6
- McBride, M.B. 1994. Trace and toxic elements in soils. In: *Environmental chemistry of the soils*. Oxford Univ. Press, p. 308–342.
- McLean, J.E., and B.E. Bledsoe. 1992. Groundwater issue (behavior of metals in soils). USEPA, Washington, DC.
- Mingkuai, Z., and W. Hao. 2009. Concentrations and chemical forms of potentially toxic metals in road-deposited sediments from different zones of Hangzhou, China. *J. Environ. Sci. (China)* 21:625–631. doi:10.1016/S1001-0742(08)62317-7
- O'Brien, R.M. 2007. A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.* 41:673–690. doi:10.1007/s11135-006-9018-6
- Park, S.-J., Z. Cheng, H. Yang, E.E. Morris, M. Sutherland, B.B.M. Gardener, and P.S. Grewal. 2010. Difference in soil chemical properties with distance to roads and age of development in urban areas. *Urban Ecosyst.* 13:483–497. doi:10.1007/s11252-010-0130-y
- Peltola, P., A. Ivask, M.   str om, and M. Virta. 2005. Lead and Cu in contaminated urban soils: Extraction with chemical reagents and bioluminescent bacteria and yeast. *Sci. Total Environ.* 350:194–203. doi:10.1016/j.scitotenv.2005.01.029
- Peng, J.-f., Y.-h. Song, P. Yuan, X.-y. Cui, and G.-l. Qiu. 2009. The remediation of heavy metals contaminated sediment. *J. Hazard. Mater.* 161:633–640. doi:10.1016/j.jhazmat.2008.04.061
- Pezzeno, M. 1997. Effects of urban environmental conditions on the symbiosis between vesicular-arbuscular mycorrhizal fungi and silver maple (*Acer saccharinum* L.). M.Sc. diss., McGill University, Montreal, Quebec, Canada.
- Pouyat, R.V., I.D. Yesilonis, J. Russell-Anelli, and N.K. Neerchal. 2007. Soil chemical and physical properties that differentiate urban land-use and cover-types. *Soil Sci. Soc. Am. J.* 71:1010–1019. doi:10.2136/sssaj2006.0164
- Quigley, M.F. 2004. Street trees and rural conspecifics: Will long-lived trees reach full size in urban conditions? *Urban Ecosyst.* 7:29–39. doi:10.1023/B:UECO.0000020170.58404.e9
- Roberts, J., N. Jackson, and M. Smith. 2006. Urban soils for amenity trees. In: *Tree roots in the built environment*. TSO, Norwich, UK.
- Sabin, L.D., J.H. Lim, K.D. Stolzenbach, and K.C. Schiff. 2005. Contribution of trace metals from atmospheric deposition to stormwater runoff in a small impervious urban catchment. *Water Res.* 39:3929–3937. doi:10.1016/j.watres.2005.07.003
- Tak ac, P., T. Szabov a, L. Koz akov a, and M. Benkov a. 2009. Heavy metals and their bioavailability from soils in the long-term polluted Central Spiš region of SR. *Plant Soil Environ.* 55:167–172.
- U.S. National Research Council Committee on Reducing Stormwater Discharge Contributions to Water Pollution. 2009. Stormwater management approaches. In: *Urban stormwater management in the United States*. National Academies Press, Washington, DC, p. 610.
- Vega, F.A., E.F. Covelo, M.L. Andrade, and P. Marcet. 2004. Relationships between heavy metals content and soil properties in minesoils. *Anal. Chim. Acta* 524:141–150. doi:10.1016/j.aca.2004.06.073
- Wei, B., and L. Yang. 2010. A review of heavy metal contaminations in urban soils, urban road dusts and agricultural soils from China. *Microchem. J.* 94:99–107. doi:10.1016/j.microc.2009.09.014
- Xia, X., X. Chen, R. Liu, and H. Liu. 2011. Heavy metals in urban soils with various types of land use in Beijing, China. *J. Hazard. Mater.* 186:2043–2050. doi:10.1016/j.jhazmat.2010.12.104
- Yobouet, Y.A., K. Adouby, A. Trokourey, and B. Yao. 2010. Cadmium, copper, lead and zinc speciation in contaminated soils. *Int. J. Eng. Sci. Technol.* 2:802–812.





# Production of polyols and mono-ols from 10 North-American vegetable oils by ozonolysis and hydrogenation: A characterization study

Marie-Josée Dumont<sup>a,\*</sup>, Ereddad Kharraz<sup>b</sup>, Hong Qi<sup>c</sup>

<sup>a</sup> Department of Bioresource Engineering, McGill University, 21 111 Lakeshore Rd., Ste-Anne-de-Bellevue, Quebec, Canada H9X 3V9

<sup>b</sup> Lipid Product Alberta, Department of Agricultural Food and Nutritional Science, 4-10 Agriculture/Forestry Centre, University of Alberta, Edmonton, Alberta, Canada T6G 2P5

<sup>c</sup> Bio-Industrial Opportunities Branch, Agri-Food Discovery Place Building, F-83, 6020 – 118 Street, Edmonton, Alberta, Canada T6H 2V8

## ARTICLE INFO

### Article history:

Received 23 April 2013

Received in revised form 2 July 2013

Accepted 4 July 2013

### Keywords:

Unrefined vegetable oils

Ozonolysis

Hydrogenation

Characterization

## ABSTRACT

Ozonolysis and hydrogenation reaction was performed on ten North-American vegetable oils. From these oils, canola oil and sunflower oil were chemically refined. Camelina, juvenile canola, three types of flax and three types of mustard were cold pressed and their oils were processed without further treatment. The production of polyols from unrefined vegetable oils by the ozonolysis–hydrogenation process is reported for the first time. The resulting polyols and mono-ols were characterized by GC-FID, DSC, GPC, HPLC, and their acid and hydroxyl numbers, viscosity and molecular weight distribution were determined. Results showed that the physical properties of the refined vegetable oils were different from the unrefined vegetable oils as they displayed higher hydroxyl values. However, the physical properties of the polyols from the unrefined starting materials are expected to be sufficient to be utilized as monomers for the production of polymeric materials.

Crown Copyright © 2013 Published by Elsevier B.V. All rights reserved.

## 1. Introduction

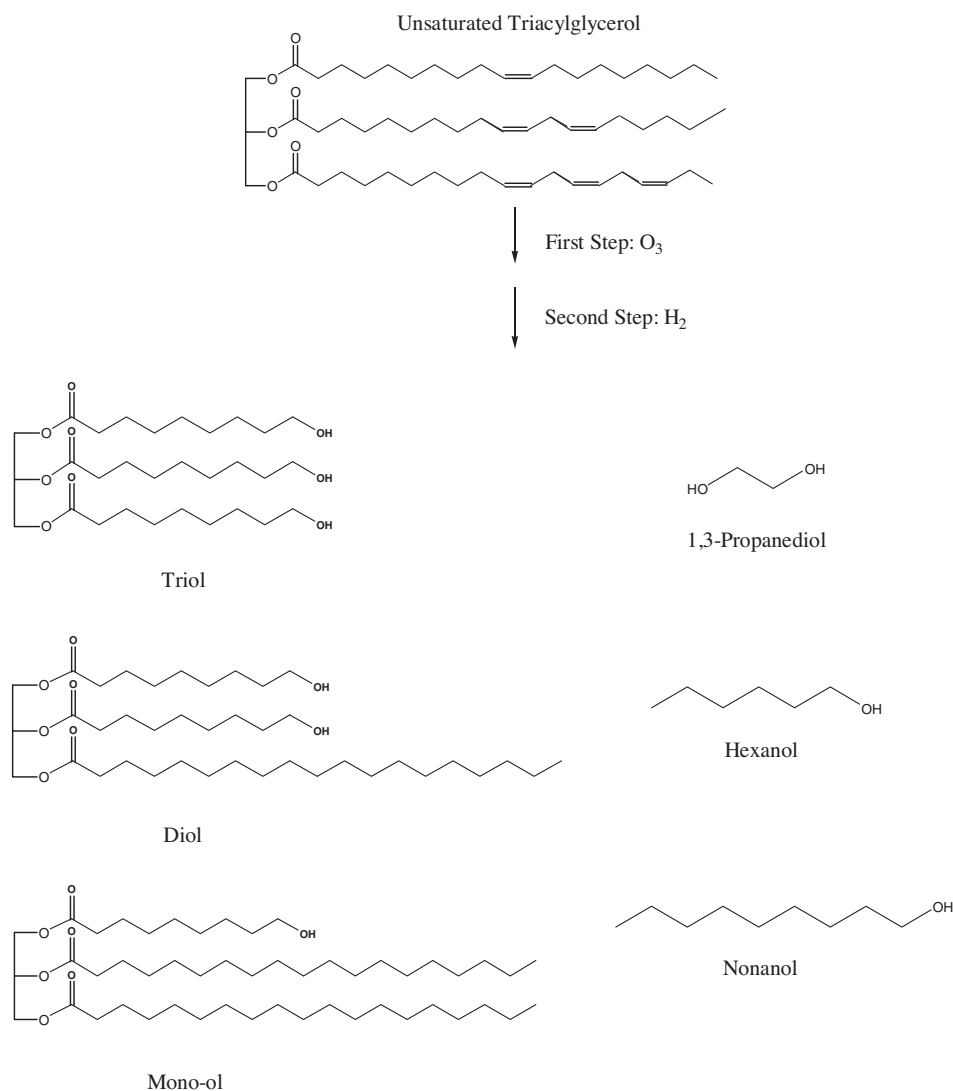
The production of polyol (poly-alcohols) monomers from vegetable oils has been an area of focus in polymer chemistry, particularly for the production of polyurethanes, to reduce the dependence on petroleum-based hydrocarbons. The epoxidation (Guo et al., 2000; Kiatsimkul et al., 2008; Kong et al., 2012) and ozonolysis–hydrogenation (Petrovic et al., 2005; Narine et al., 2007a) reactions are efficient and well-known chemical routes to introduce hydroxyl groups on triacylglycerols (TAGs). There are major structural differences found in the resulting monomers depending on the chemical route favored. Through epoxidation, the hydroxyl groups are located in a secondary position while through ozonolysis–hydrogenation, the hydroxyl groups are in a primary position. Furthermore, the co-products of the latter reaction are of commercial interest. These are short-chain mono-alcohols (mono-ols) or di-alcohols (diols) (Fig. 1). Mono-ols, such as hexanol, are used by the perfume industry (Kandra and Wagner, 1998) whereas diols, such as 1,3-propanediol, are used as chain extenders (Miao et al., 2013; Rashmi et al., 2013). The main drawbacks of the ozonolysis–hydrogenation reaction are the formation of acid at

the ozonolysis stage and the high cost of the reaction. Recently, Omonov et al. (2011) focused their efforts to optimize the ozonolysis step by selecting conditions in order to reduce the amount of carboxylic acid formed in the aldehyde pool. It is important to note that the concentration of acid in the resulting polyols is not above the limit to prevent the formation of polyurethanes (Petrovic et al., 2005; Kong et al., 2007; Narine et al., 2007b,c).

This publication is focused on the production of polyols, diol and mono-ols from unrefined vegetable oils in order to reduce the cost and increase the environmental friendliness of the ozonolysis–hydrogenation process. The unrefined vegetable oils selected were mustard oils (brown, yellow and oriental), flax oils (Nulin 50, CDC Bethune and Linola 2090), camelina oil and immature canola oil. The flaxes are of particular interest since they were genetically modified and therefore contained a unique fatty acid profile. Mustard oils were selected due to their high content in erucic acid which renders them indigestible for humans (Al-Jasass and Al-Jasser, 2012). Camelina oil was selected since it is used as industrial oil (Zubr, 1997) and immature canola was selected since it does not meet the allowable chlorophyll limit for food consumption (Daun and Symons, 2000; Daun, 2003). Because unrefined vegetable oils were not physically or chemically treated, they contained molecules other than TAGs such as sterols and tocopherols. Therefore, the study of the influence of the non-TAG content on the resulting polyols, mono-ols and diols was an objective of this study.

\* Corresponding author. Tel.: +1 514 398 7776; fax: +1 514 398 8387.

E-mail address: [marie-josee.dumont@mcgill.ca](mailto:marie-josee.dumont@mcgill.ca) (M.-J. Dumont).



**Fig. 1.** Scheme of ozonolysis–hydrogenation reaction of vegetable oils.

The polyols and light alcohol content of these oils were compared to canola oil and sunflower oil which acted as standards since these oils were chemically refined.

## 2. Materials and methods

### 2.1. Chemicals and materials

Refined sunflower oil was supplied by Bunge oil (Alberta), canola oil was supplied by Cargill Canada (Alberta), seeds of flax (Nulin 50, CDC Bethune and Linola 2090) were supplied by Viterra Inc. Vegreville (Alberta), and seeds from brown, oriental, and yellow mustard were supplied by Viterra Specialty Crops (Alberta), cold pressed immature canola oil and seeds of camelina were supplied by local farmers (Alberta).

Ethyl acetate (ACS reagent  $\geq 99.5\%$ ) was purchased from Sigma–Aldrich and used as solvent for ozonolysis reaction. Sponge Nickel (A-7063) promoted with 1% molybdenum (slurry in water) was obtained from Alfa Aesar and used as the hydrogenation catalyst. Celite was obtained from Fisher Scientific USA.

The following reagents were used as calibrating standards for GC-FID. Propanal ( $\geq 98\%$ ), propanoic acid ( $\geq 98\%$ ), hexanal ( $\geq 98\%$ ), nonanal ( $\geq 98\%$ ), methyl nonanoate ( $\geq 98\%$ ), ethyl nonanoate ( $\geq 98\%$ )

and heptanoic acid ( $\geq 99\%$ ) were purchased from Sigma–Aldrich. Propanol (99%), hexanol (99%), heptanol (99%), nonanol (97%), hexanoic acid (99%), nonanoic acid (97%) were purchased from Acros. 1,3-Propanediol ( $\geq 99\%$ ) was supplied by DuPont Tate & Lyle Zemea. p-Anisaldehyde ( $\geq 98\%$ ) was purchased from Sigma–Aldrich and used as an internal standard for the GC-FID measurements.

### 2.2. Cold press

A Komet screw oil expeller was used to crush the seeds of the different varieties of flax and mustard in order to extract the vegetable oil. The processing rate ranged from 4 to 6 kg/h.

### 2.3. Centrifugation

Cold press mustard, camelina, flax and juvenile canola oils were centrifuged with a Beckman Coulter Avanti J26XPI centrifuge with a JLA 8.1000 rotor. The speed used was 6500 rpm (which equals  $10,543 \times g$ ) for 30 min at room temperature.

### 2.4. Vegetable oil characterization

The analysis of fatty acid methyl esters (FAMES) was carried out using an Perkin Elmer Clarus 500 gas chromatograph

(Perkin Elmer Instruments LLC, Shelton, CT) equipped with a flame-ionization detector (GC-FID) and an SP-2560 capillary column (100 m × 0.25 mm i.d., 0.2 μm film thickness, Supelco). The oven temperature was maintained at 45 °C for 4 min, increased to 175 °C at a rate of 13 °C/min, maintained at 175 °C for 27 min then increased to 215 °C at a rate of 4 °C/min, and finally held at 215 °C for 35 min. The injector and detector temperatures were 240 °C and 280 °C, respectively. Hydrogen was used as carrier gas at a rate of 1 mL/min and the injection port was in split mode of 40:1.

The vegetable oil samples (≈40 mg) were dissolved in toluene (1 mL) in a Pyrex capped test tube where 2 mL of 10% of acetyl chloride in methanol were added. The test tube was sealed and placed in an oven at 80 °C for 2 h. 5 mL of water containing 5% of sodium chloride was added to the tube after cooled down to room temperature. The formed esters were extracted with (2 × 5 mL) of hexane. The hexane layer was washed with basic water (2% sodium bicarbonate) and dried over anhydrous sodium sulphate. The hexane and sodium sulphate were centrifuged for 2 min. C17:1 (internal standard) was added to the 5 mL hexane layer for GC injection.

The individual FAMEs were identified by their retention times, and by comparison with FAME reference standards (GLC-463, Nu-chekprep). All concentrations were calculated by using methyl 10-heptadecenoate (C17:1) as internal standard. The AOCS official Method Ce 1h-05 was applied to determine the theoretical flame ionization detector correction factor. The samples were run in duplicate.

### 2.5. Ozonolysis

Vegetable oils (100 g) and ethyl acetate (600 mL) were added to an ozonolysis batch reactor as described in Omonov et al. (2011). The reaction was carried out at −10 °C under an ozone concentration of 6.5 L/min. The agitation was set at a constant speed (500 rpm) and the duration of the reaction was of 45 min. After the reaction, the unreacted ozone was purged with nitrogen gas for 5 min. The ozonolysis product (aldehyde oil) was sampled for GC analysis and then transferred for hydrogenation. The samples were run in duplicate.

### 2.6. Hydrogenation

The aldehyde oil was transferred to a hydrogenation vessel described in Kong and Narine (Narine et al., 2007a). Then 10 g of nickel catalyst (Sponge Nickel, A-7063, promoted with 1% molybdenum) were added. The reaction vessel was charged with hydrogen gas to a pressure of 300 psi. The temperature was then increased to 100 °C. The reaction was carried out for 3 h prior for the temperature to be reduced to room temperature. The reaction vessel was finally purged with nitrogen gas and the resulting hydrogenated mixture was filtered over celite. The solvent was then removed from the resulting mixture using rotary evaporation. The samples were run in duplicate.

### 2.7. Wipped-blade molecular distillation

The final hydrogenation product was added to a wipped-blade molecular distillation vessel (Model VKL 70/ICL-04, from Incon Processing, IL, USA) to separate the lighter and heavier alcohols. The jacket temperature was maintained at 115 °C, the condenser was maintained at 5 °C and the vacuum was maintained in the range of 80–90 mTorr.

### 2.8. Characterization of the reaction products

The GC apparatus was used for the characterization of the aldehyde (ozonolysis product) and the low molecular weight alcohols

recovered from the distillation step. It was also used as a tool for the determination of the hydrogenation completion. The apparatus was an Agilent 6890. A split injection 80:1 was used for 1 μL of sample. Helium was the carrier gas and the flow rate was 1.5 mL/min. The initial oven temperature was held at 50 °C for 3 min, rising to 90 °C at 25 °C/min and then to 250 °C at 10 °C/min and held for 13 min. The universal flame ionization detector had a range of 0–10 V. A BP20 chromatographic column (30 m × 0.25 mm i.d. × 0.25 μm film thickness) (SGE, USA) was selected. The detection of 1,6-hexanediol was done similarly except that a 5975B MSD Electron impact mode was used.

The alcohol and acid standard solutions were prepared by dissolving the standards (listed in Section 2.1) in dichloromethane. All the standard solutions had a concentration of 10 mg mL<sup>−1</sup> and were injected with the internal standard. Standard calibration curves were generated by the analysis of a series of standard solutions over appropriate concentration ranges. The fatty acid standards have been injected within the concentration range of 0.015 mg mL<sup>−1</sup> and 0.38 mg mL<sup>−1</sup>. All calibration concentration points were run in triplicate.

### 2.9. Hydroxyl and acidity values

The hydroxyl number of the polyols was determined according to ASTM D1957-86 and the acidity values were determined according to the ASTM D4662-98.

### 2.10. Rheometric measurements

The viscosity of the polyols monomers was measured with an AR 2000 Rheometer (TA Instrument, Delaware) in a shearing mode with a shearing rate of 50.0 s<sup>−1</sup>.

### 2.11. Molecular weight distribution

The molecular weight distribution of the polyols was determined with a gel permeation chromatography high pressure liquid chromatography (GPC HPLC) model 1200 series from Agilent equipped with an evaporative light scattering detector (ELSD) 2000 from Alltech. Polyols (10 mg) were dissolved in 10 mL of dichloromethane. Then, 50% (v/v) of the mother solution and dichloromethane were transferred into the injection vial. The injection volume was 10 μL. THF was used as the mobile phase and the flow rate was of 1 mL/min. The temperature of the ELSD detector was 70 °C. The gas flow rate and gain were 2 L/min and 1 respectively. A Styragel HR 1 THF chromatographic column (7.8 mm × 300 mm) (Waters, USA) was selected. High molecular weight (Mw) polystyrenes (Mw: 10,200, 6520, 3950, 2340) from Shodex STANDARD (Japan), tripalmitic triglycerol (PPP) from Sigma-Aldrich and distearine and monostearine from Nu-chekprep (USA) were used as standards.

### 2.12. HPLC

The semi-quantitative analysis of the mono-ols, diols and tri-ols fraction was done with an HPLC model 1200 series from Agilent equipped with an ELSD 2000 detector from Alltech. Polyols (10 mg) were dissolved in 10 mL of dichloromethane. Then, 50% (v/v) of the mother solution and dichloromethane were transferred into the injection vial. The injection volume was 10 μL of samples. Hexane:isopropanol (1:1) and hexane (100%) were used as the gradient mobile phase. The flow rate was 3 mL/min and the gradient started from 100% to 70% hexane in 20 min. The temperature of the ELSD detector was 100 °C. The gas flow rate and gain were 2 L/min and 8 respectively. The column temperature was set at 50 °C. A Thermo

**Table 1**  
Fatty acid composition (% mass) of vegetable oils.

Fatty acids	Canola	Juvenile canola	Sunflower	Camelina	Oriental mustard	Brown mustard	Yellow mustard	Nulin 50 Flax	CDC Bethune Flax	Linola 2090 Flax
	Mass content (%) <sup>a</sup>									
Myristic C14:0	0.060	0.070	0.090	0.060	0.070	0.070	0.070	0.050	0.040	0.050
Palmitic C16:0	4.1	4.9	6.4	5.6	3.1	3.40	3.2	4.9	5.1	5.3
Palmitoleic 16:1(n-7)	0.24	0.25	0.10	0.10	0.14	0.19	0.15	0.040	0.050	0.050
Stearic C18:0	1.8	1.8	3.4	2.3	1.5	1.3	1.1	2.4	3.8	3.9
Oleic 18:1(n-9)	66	54	30	16	19	17	28	9.0	20	17
Vaccenic 18:1(n-7)	3.4	3.9	0.90	0.94	1.4	1.9	1.2	0.57	0.63	0.55
Linoleic 18:2(n-6)	19.4	21	57	21	21	22	9.1	13	16	71
$\alpha$ -Linolenic 18:3(n-3)	2.50	12	0.47	33	13	15	11	70	54	1.7
Arachidic C20:0	0.62	0.65	0.26	1.2	0.85	0.79	0.63	0.060	0.12	0.10
Gondoic 20:1(n-9)	1.2	1.3	0.24	13	12	11	10	0.00	0.10	0.15
Eicosadienoic 20:2(n-6)	0.050	0.090	0.47	1.8	1.1	1.3	0.30	0.030	0.020	0.070
Eicosatrienoic 20:3(n-3)	0.00	0.00	0.00	1.0	0.00	0.00	0.00	0.10	0.040	0.00
Behenic C22:0	0.34	0.42	0.62	0.26	0.45	0.40	0.47	0.080	0.12	0.14
Erucic 22:1(n-9)	0.00	0.00	0.00	2.4	24	22	32	0.00	0.00	0.00
Docosadienoic 22:2(n-6)	0.00	0.00	0.00	0.26	0.44	0.39	0.47	0.080	0.12	0.14
Lignoceric C24:0	0.17	0.00	0.22	0.13	0.25	0.29	0.25	0.050	0.080	0.10
Nervonic 24:1(n-9)	0.15	0.21	0.00	0.89	1.6	1.8	2.1	0.00	0.00	0.00
Saturated	7.1	7.8	11	9.6	6.2	6.2	5.7	7.5	9.3	9.6
Mono-unsaturated	71	60	31	33	58	54	73	9.6	21	18
Di-unsaturated	19	21	57	23	23	23	9.9	13	16	71
Tri-unsaturated	2.5	12	0.47	34	13	15	11	70	54	1.7

<sup>a</sup> Maximum standard deviation:  $\pm 0.7$ .

Fisher Betasil diol-100 chromatographic column (particle size of 5  $\mu\text{m}$ , 4 mm  $\times$  250 mm) (USA) was selected.

### 2.13. Differential scanning calorimetry

A differential scanning calorimeter (DSC) Q100 from TA Instruments (DE, USA) was used for the DSC measurements. This apparatus was equipped with a refrigerated cooling system and the measurements were performed under a nitrogen gas atmosphere. The measurements were divided into 2 cycles. Cycle one consisted in cooling the sample (about 7 mg) at a rate of 3 °C/min from room temperature to  $-15$  °C, holding the temperature for 5 min. Cycle two consisted in heating the samples to 70 °C at a rate of 5 °C/min. The measurements were done in duplicate.

## 3. Results and discussion

### 3.1. Vegetable oils characterization

Table 1 shows the fatty acid composition of canola oil, juvenile canola oil, sunflower oil, camelina oil, three types of mustard oils and three types of flax oils. The canola and sunflower oils were received in a chemically refined state and hence were mainly composed of triacylglycerol. Seeds of mustards, flaxes and camelina were cold pressed to extract the oils. The oils were subsequently centrifuged. Because no refining treatments were performed on these oils, different classes of lipid materials such as free fatty acids, phospholipids, sterols and tocopherols (Dumont and Narine, 2007) were present in the bulk. Furthermore, juvenile canola oil is known for its high concentration in chlorophyll (above 200 mg/kg) (Daun and Symons, 2000; Daun, 2003), which may also impact the reaction products from the ozonolysis–hydrogenation process. The juvenile canola oil was also centrifuged prior to its characterization.

From Table 1, the mono-, di- and tri-unsaturated fatty acid profile varied greatly among the species and among the varieties of the same species while the unsaturated fatty acid content did not vary significantly. The canola and the juvenile canola oils contained oleic acid as the main fatty acid. The main fatty acid of the oriental and yellow mustard oils was erucic acid. Brown mustard oil had the same proportion (22 wt.%) of linoleic and erucic acid. The main fatty acid found in sunflower and camelina oils was linoleic acid.

To our best knowledge, the fatty acid composition of CDC Bethune, Nulin 50 and Linola 2090 is reported for the first time. The fatty acid composition of these flax varieties showed great discrepancies. From Table 1, Nulin 50 is composed of 70 wt.% of tri-unsaturated fatty acids which correspond to its content in  $\alpha$ -linolenic acid. The CDC Bethune variety holds the second highest content in  $\alpha$ -linolenic acid, with a ratio of 54 wt.%. Interestingly, the proportion of tri-unsaturated fatty acid of the Linola 2090 variety is very low (1.7 wt.%) but holds the highest percentage of di-unsaturated fatty acid (71 wt.%) due to its content in linoleic acid.

### 3.2. Aldehyde formation

It is well known that ozonolysis of fatty acids leads to the formation of aldehydes. The efficiency of the ozonolysis step was monitored through quantification of nonanal; the aldehyde formed by reacting oleic acid with ozone. The choice of nonanal was based on the fatty acid profile of the vegetable oils, which showed that oleic acid was significantly present in the oil samples (Table 2). From Table 1, high oleic canola and juvenile canola oils held the highest concentration with  $\approx 66\%$  and  $\approx 54\%$  respectively. Nulin 50 held the lowest concentration with  $\approx 9\%$ . Predictably, there was a correlation between the nonanal content of the ozonide and the oleic acid content of the different vegetable oils while considering the standard deviations (Table 2). High standard deviations were also reported for similar measurements (Omonov et al., 2011).

**Table 2**  
Nonanal content (mol/100 g oil) after ozonolysis of different vegetable oils.

Vegetable oil	Nonanal content (mol/100 g oil)
High oleic canola refined	0.062 $\pm$ 15%
Juvenile canola	0.060 $\pm$ 14%
Sunflower oil refined	0.024 $\pm$ 19%
Camelina	0.027 $\pm$ 6%
Oriental mustard	0.049 $\pm$ 9%
Brown mustard	0.043 $\pm$ 19%
Yellow mustard	0.063 $\pm$ 4%
Nulin 50	0.0049 $\pm$ 5%
CDC Bethune	0.014 $\pm$ 7%
Linola 2090	0.014 $\pm$ 9%



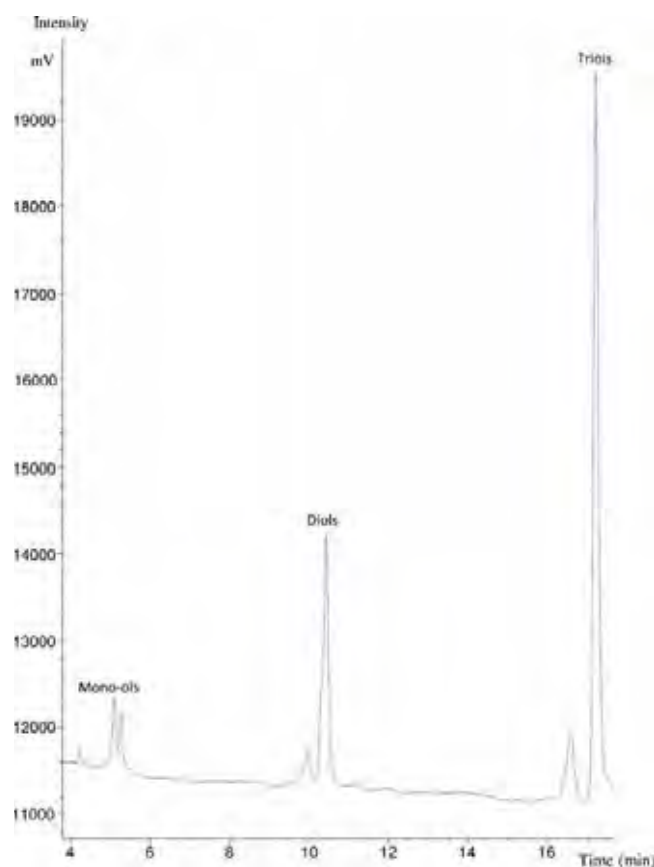


Fig. 2. Semi-quantitative distribution of triols, diols and mono-ols of juvenile canola oil.

### 3.3. Physical properties of polyols

As pictured in Fig. 1, the hydrogenation of the products yielded from the ozonolysis step (the ozonide) led to the formation of high molecular weight polyols and light alcohols. It was measured by GC-FID that the hydrogenation was completed at  $\geq 99\%$  for all samples. After hydrogenation, the light alcohols were separated from the higher molecular weight polyols through distillation. The semi-quantitative distribution of triols, diols and mono-ols was performed by normal phase HPLC. Fig. 2 shows the juvenile canola polyol fraction distribution which is similar to the polyols from the 9 other vegetable oils. Table 3 shows that the triols are generally the main component of the polyols. The proportion of triols in canola, juvenile canola, sunflower and mustard polyols accounted for at least 62 wt.%. The triol concentration in camelina and flax polyols accounted for less than 50%. The results showed that camelina and Nulin 50 polyols had a higher concentration in diols than triols.

Table 3  
Polyols fraction distribution.

	Saturated TAGs	Mono-ols	Diols	Triols
Canola (refined)	0	6% $\pm$ 1	11% $\pm$ 1	83% $\pm$ 2
Juvenile Canola	0	7% $\pm$ 0.6	30% $\pm$ 6	62% $\pm$ 6
Sunflower oil (refined)	0	6% $\pm$ 0.5	28% $\pm$ 3	65% $\pm$ 3
Camelina	3% $\pm$ 0.4	26% $\pm$ 1	44% $\pm$ 1	27% $\pm$ 1
Oriental mustard	1% $\pm$ 0.2	7% $\pm$ 0.5	27% $\pm$ 3	65% $\pm$ 3
Brown mustard	1% $\pm$ 0.1	11% $\pm$ 2	26% $\pm$ 2	62% $\pm$ 2
Yellow mustard	0	3% $\pm$ 1	19% $\pm$ 2	77% $\pm$ 2
Nulin 50	6% $\pm$ 2	29% $\pm$ 2	35% $\pm$ 3	30% $\pm$ 2
CDC Bethune	4% $\pm$ 1	15% $\pm$ 4	38% $\pm$ 6	43% $\pm$ 10
Linola 2090	2% $\pm$ 1	11% $\pm$ 3	38% $\pm$ 3	49% $\pm$ 7

Therefore, there seemed to be a correlation between the triol concentration and the hydroxyl number of the polyols (Table 4), where camelina and flax polyols displayed the lowest hydroxyl numbers. As shown in Table 1, camelina, CDC Bethune and Nulin 50 had the highest concentration in linolenic acid (a tri-unsaturated fatty acid) which may have required a longer ozonolysis time for conversion. Similarly, Linola 2090 had the highest concentration in linoleic acid (a di-unsaturated fatty acid). This may explain the low concentration in triols and lower hydroxyl numbers. However, it is important to note that Nulin 50, which displays the highest saturated Tag content (6%) and mono-ol content (29%) could still be utilized as polyol for the production of polyurethanes. It was shown that vegetable oil-based polyols with a mono-ol and saturated TAGs content of 26% and 14% respectively were used for that purpose (Kong et al., 2007).

The difference between the calculated and the experimental hydroxyl number is of higher importance for the polyols synthesized from unrefined vegetable oils. Canola and sunflower oils had a low concentration in tri-unsaturated fatty acids. Furthermore, because these oils were refined, no impurities such as sterols and tocopherols were present in the lipidic pool. Therefore, their higher hydroxyl numbers were expected.

The hydroxyl numbers of the flax species are especially low. Linola 2090 had a higher concentration in triols and a higher hydroxyl number than the other two mustard species, probably because of its low concentration in tri-unsaturated fatty acid. Nulin 50 and CDC Bethune held the highest concentration in tri-unsaturated fatty acids. Linola 2090 also has a very low concentration in tri-unsaturated fatty acids but holds the highest concentration in di-unsaturated fatty acids, which may explain the highest hydroxyl number compared to the other flax varieties. The low hydroxyl number of camelina was unexpected. Because camelina has a similar content in mono-, di- and tri-unsaturated fatty acids as brown and oriental mustard, camelina's hydroxyl number was expected to be similar to these mustard species. Because the oils were not refined prior to their processing, these oils contained non-saponifiable matter. Therefore, the lower hydroxyl number found for camelina may be the result of a higher content in non-saponifiable matter which may have consumed part of the available hydrogen destined for the hydrogenation of the ozonide and therefore, led to the resulting camelina polyol with a lower hydroxyl number. However, further investigation is needed.

Nonetheless, the experimental hydroxyl number followed the trend of the calculated hydroxyl number. For example, the mustard polyol family had a higher hydroxyl number than the flax polyol family. However, the trend is not linear within the same oil family. Table 4 shows that Linola 2090 has a higher calculated hydroxyl number but a lower experimental hydroxyl number than Nulin 50 and CDC Bethune. In this case, Nulin 50 only contained 17% of tri-unsaturated fatty acid as compared to Linola 2090 and CDC Bethune which contained 70% and 54% of tri-unsaturated, respectively.

The number average molecular weight of the polyols sourced from refined vegetable oils is also lower than sourced from crude vegetable oils (Table 4). This is explained by the non-triacylglycerol content present in crude oils. For example, the polyols from juvenile canola were 60 g/mol heavier than the polyols from refined canola oil even though canola and juvenile canola oils displayed similar fatty acid profiles. The main difference between the juvenile and the refined canola oils is the ratio of oleic and linolenic acid, which are both fatty acids containing 18 carbons. Therefore the presence of chlorophyll and other compounds such as sterols and tocopherols are responsible for the difference in weight.

Both the number average molecular weight and the hydroxyl number influence the viscosity of the polyols (Fig. 3). Therefore, various levels of crystallinity are present. For example, polyols from mustard (oriental, brown and yellow mustard) were crystallizing



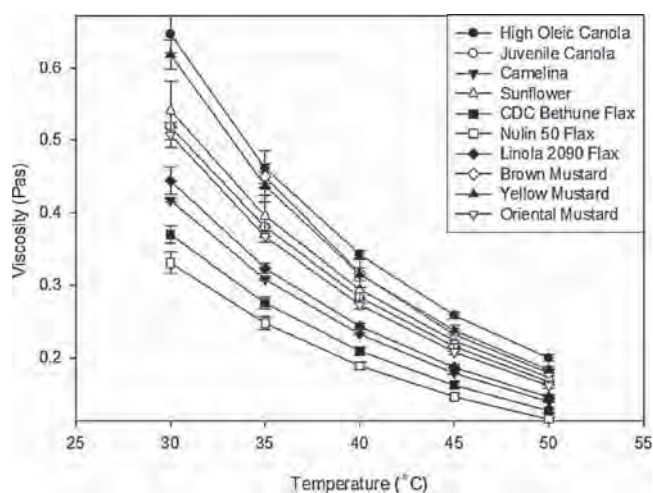


Fig. 3. Viscosity (Pa.s) of polyols in function of the temperature.

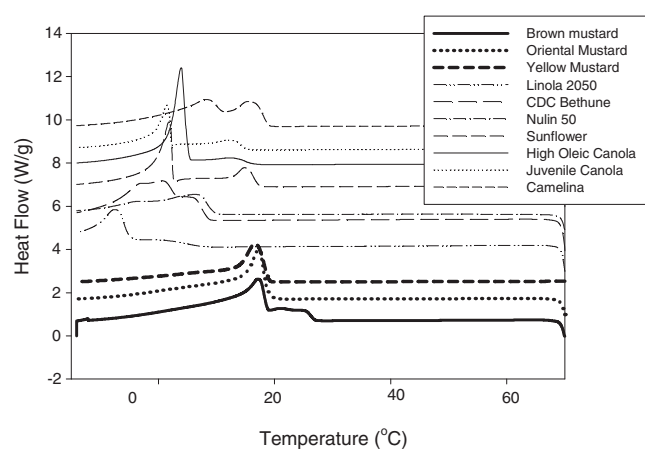


Fig. 4. Crystallization pattern of polyols.

before 15 °C (Fig. 4) and hence was the most viscous material. This is not surprising since the materials had high number average molecular weights (666–692 g/mol) and high hydroxyl numbers (184–195 mg KOH/g). On the contrary, the lowest onset of crystallization is found for all flax species (about –7 °C). The hydroxyl number of Nulin 50 and CDC Bethune are amongst the lowest (165 and 163 mg KOH/g respectively). Linola 2090 has a high hydroxyl number (180 mg KOH/g) but the number average molecular weight is low (612 g/mol). This trend is also demonstrated by the viscosity profile of the polyols since all polyols synthesized from the flax species show lower viscosities than polyols synthesized from the mustard species (Fig. 3). The highest viscosity is found for the refined canola polyols, which display a hydroxyl number above

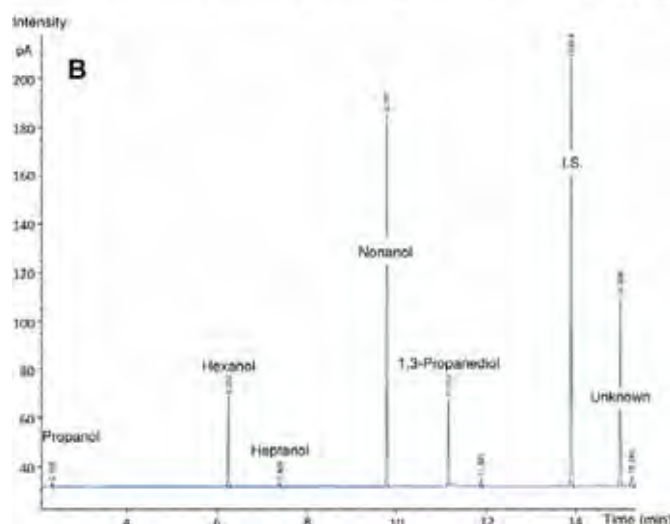
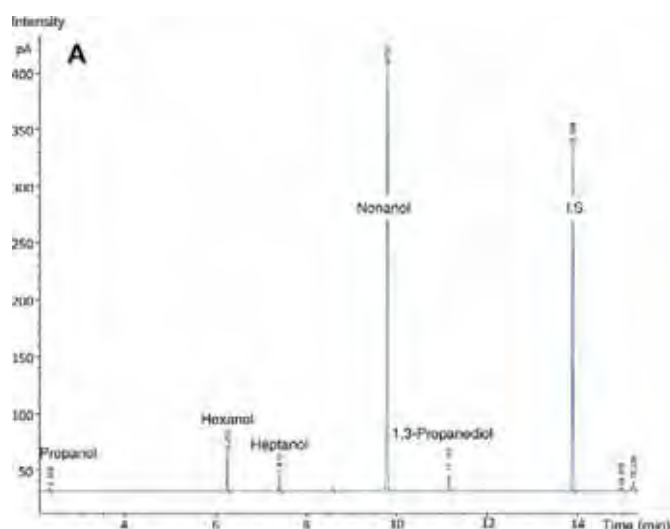


Fig. 5. (A) Light alcohol fraction from canola oil (refined). (B) Light alcohol fraction from juvenile canola oil.

200 mg KOH/g but a low molecular weight value of 521 g/mol. The increase in OH groups increases the chance of oligomer formation (Petrovic, 2008) which would increase the viscosity.

#### 3.4. Characterization of light fraction

Fig. 5a and b presents the light alcohols resulting from the ozonolysis–hydrogenation of canola and juvenile canola oil, respectively. As suggested by the figures, nonanol is the main component in the bulk. Both figures show similar light alcohol distribution

Table 4

Hydroxyl number (mg KOH/g), acid number (mg KOH/g) and number average molecular weight of polyols from different vegetable oils.

Polyol source	Hydroxyl number (mg KOH/g)	Calculated hydroxyl number (mg KOH/g)	Acid number (mg KOH/g)	Number average molecular weight (g/mol)
Canola (refined)	203 ± 6%	278	11 ± 2%	521 ± 0.8%
Juvenile canola	185 ± 9%	276	11 ± 4%	581 ± 0.7%
Sunflower oil (refined)	210 ± 9%	266	11 ± 12%	563 ± 4%
Camelina	165 ± 4%	270	7 ± 5%	692 ± 0.6%
Oriental mustard	194 ± 3%	280	9 ± 6%	666 ± 0.5%
Brown mustard	184 ± 3%	280	9 ± 3%	683 ± 0.4%
Yellow mustard	195 ± 2%	282	11 ± 7%	654 ± 1%
Nulin 50	165 ± 3%	276	8 ± 3%	699 ± 2%
CDC Bethune	163 ± 7%	271	8 ± 3%	660 ± 3%
Linola 2090	180 ± 5%	270	10 ± 13%	612 ± 2%

**Table 5**

Light alcohol content (g/L of vegetable oil) in function of the vegetable oil source.

	Hexanol (g/L)	Heptanol (g/L)	Nonanol (g/L)	1,3-Propanediol (g/L)
Canola (refined)	2.7 ± 13%	1.6 ± 14%	47 ± 11%	6.1 ± 6%
Juvenile canola	3.3 ± 8%	2.2 ± 6%	50.3 ± 5%	7.2 ± 10%
Sunflower oil (refined)	6.6 ± 10%	0.2 ± 15%	35.7 ± 19%	15.5 ± 11%
Camelina	3.8 ± 25%	0.4 ± 23%	37.4 ± 4%	10.9 ± 4%
Oriental mustard	3.6 ± 11%	1.1 ± 15%	46.5 ± 10%	7.9 ± 6%
Brown mustard	3.7 ± 8%	1.5 ± 12%	42.8 ± 9%	9.3 ± 13%
Yellow mustard	1.7 ± 6%	0.9 ± 4%	46.0 ± 3%	6.0 ± 6%
Nulin 50	4.4 ± 12%	0	25.9 ± 13%	18.7 ± 11%
CDC Bethune	7.9 ± 4%	0	25.4 ± 6%	13.2 ± 8%
Linola 2090	7.1 ± 8%	0	30.3 ± 17%	13.2 ± 16%

patterns. However, an unknown peak at approximately 15 min is present in the juvenile canola light fraction. It was determined that the mass and the probable structure of the fraction to be similar to hexanediol (molecular weight = 118). This may be due to the formation of oligomers as suggested by Omonov et al. (2011).

Table 5 presents the light alcohol fraction for all vegetable oils. Nonanol is the main component for all fractions. This is due to the oleic acid content in vegetable oils but also because erucic acid (*n-9*) is also a main component in mustard oils (Table 1).

#### 4. Conclusion

This study showed that it is possible to produce polyols and light alcohols using crude vegetable oils as feedstock. It was demonstrated that the physical properties of polyols from unrefined vegetable oils differ from the refined vegetable oils. The presence of high molecular weight compounds such as phospholipids, sterols, tocopherols, chlorophyll and others increase the number average molecular weight of the polyol pool. It was also found that the hydroxyl numbers were lower than when refined vegetable oils are used as feedstock. Even though polyols from crude vegetable oils exhibited substantial gaps between the calculated and experimental hydroxyl number, some examples from the literature showed that polyurethanes elastomers and foams could be synthesized within the hydroxyl number range obtained in this study (Narine et al., 2007a,c). Our future work will provide the physical properties of polyurethane matrices from polyols synthesized from unrefined vegetable oils.

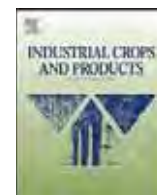
The characterization of the light fraction revealed a good correlation with the fatty acid profile. The preponderance of nonanol in the bulk was due to oleic acid and eicosanoic acid, with both being found in large quantities in the vegetable oils.

#### Acknowledgments

The financial support of the Alberta Crop Industry Development Fund and the Alberta Canola Producers Commission are greatly acknowledged. We would also like to thank Mr. Niranjan Purohit for his technical expertise.

#### References

- Al-Jasass, F.M., Al-Jasser, M.S., 2012. Chemical composition and fatty acid content of some spices and herbs under Saudi Arabia conditions. *Scientific World Journal* 2012, 859892.
- Daun, J.K., 2003. How green is green? Long-term relationships between green seeds and chlorophyll in canola grading. *J. Am. Oil Chem. Soc.* 80, 119–122.
- Daun, J.K., Symons, S., 2000. How green is green? Sampling and perception in assessing green seeds and chlorophyll in canola. *J. Am. Oil Chem. Soc.* 77, 1209–1213.
- Dumont, M.J., Narine, S.S., 2007. Characterization of flax and soybean soapstocks, and soybean deodorizer distillate by GC-FID. *J. Am. Oil Chem. Soc.* 84, 1101–1105.
- Guo, A., Cho, Y.J., Petrovic, Z.S., 2000. Structure and properties of halogenated and nonhalogenated soy-based polyols. *J. Polym. Sci. A: Polym. Chem.* 38, 3900–3910.
- Kandra, L., Wagner, G.J., 1998. Pathway for the biosynthesis of 4-methyl-1-hexanol volatilized from petal tissue of *Nicotiana sylvestris*. *Phytochemistry* 49, 1599–1604.
- Kiatsimkul, P.P., Suppes, G.J., Hsieh, F.H., Lozada, Z., Tu, Y.C., 2008. Preparation of high hydroxyl equivalent weight polyols from vegetable oils. *Ind. Crops Prod.* 27, 257–264.
- Kong, X.H., Liu, G.G., Curtis, J.M., 2012. Novel polyurethane produced from canola oil based poly(ether ester) polyols: synthesis, characterization and properties. *Eur. Polym. J.* 48, 2097–2106.
- Kong, X.H., Yue, J., Narine, S.S., 2007. Physical properties of canola oil based polyurethane networks. *Biomacromolecules* 8, 3584–3589.
- Miao, S.D., Zhang, S.P., Su, Z.G., Wang, P., 2013. Synthesis of bio-based polyurethanes from epoxidized soybean oil and isopropanolamine. *J. Appl. Polym. Sci.* 127, 1929–1936.
- Narine, S.S., Yue, J., Kong, X.H., 2007a. Production of polyols from canola oil and their chemical identification and physical properties. *J. Am. Oil Chem. Soc.* 84, 173–179.
- Narine, S.S., Kong, X.H., Bouzidi, L., Sporns, P., 2007b. Physical properties of polyurethanes produced from polyols from seed oils: I. Elastomers. *J. Am. Oil Chem. Soc.* 84, 55–63.
- Narine, S.S., Kong, X.H., Bouzidi, L., Sporns, P., 2007c. Physical properties of polyurethanes produced from polyols from seed oils: II. Foams. *J. Am. Oil Chem. Soc.* 84, 65–72.
- Omonov, T.S., Kharraz, E., Curtis, J.M., 2011. Ozonolysis of canola oil: a study of product yields and ozonolysis kinetics in different solvent systems. *J. Am. Oil Chem. Soc.* 88, 689–705.
- Petrovic, Z.S., 2008. Polyurethanes from vegetable oils. *Polym. Rev.* 48, 109–155.
- Petrovic, Z.S., Zhang, W., Javni, I., 2005. Structure and properties of polyurethanes prepared from triglyceride polyols by ozonolysis. *Biomacromolecules* 6, 713–719.
- Rashmi, B.J., Rusu, D., Prashantha, K., Lacrampe, M.F., Krawczak, P., 2013. Development of water-blown bio-based thermoplastic polyurethane foams using bio-derived chain extender. *J. Appl. Polym. Sci.* 128, 292–303.
- Zubr, J., 1997. Oil-seed crop: *Camelina sativa*. *Ind. Crops Prod.* 6, 113–119.



# Processing and physical properties of canola protein isolate-based films



Weida Shi, Marie-Josée Dumont\*

Bioresource Engineering Department, McGill University, 21111 Lakeshore Rd., Ste-Anne-de-Bellevue, QC, Canada H9X 3V9

## ARTICLE INFO

### Article history:

Received 24 July 2013

Received in revised form 17 October 2013

Accepted 21 October 2013

### Keywords:

Canola protein isolate

Casting film

Mechanical properties

Thermal properties

Morphological properties

Water absorption

## ABSTRACT

The utilization of canola proteins as feedstock for bio-based polymers has not been demonstrated extensively. However, canola proteins has an amino acid profile similar to that of soy proteins, suggesting that canola protein may be a potential feedstock for manufacturing biodegradable polymeric materials. This study reports for the first time, the synthesis of canola protein isolate-based films plasticized with glycerol by solution casting. Sodium dodecyl sulfate (SDS) and stearic acid (SA) were used as the protein denaturant and co-plasticizer respectively. The functional properties of the films, including the structural, mechanical, thermal, water absorptive and morphological properties were determined. It was found that SDS substantially increased the mechanical properties and raised the water absorption capacities of the resulting film up to 1115%. However, films co-plasticized by SA had weaker mechanical properties, and lower water absorption capacities.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

Canola was bred from rapeseed and it differentiates from the latter by its low content in erucic acid; an anti-nutritive fatty acid which compromised the value of rapeseed oil for years (Shahidi, 1990; Tan et al., 2011). The successful reduction in erucic acid combined with a low level of saturated fatty acids makes canola oil one of the most popular edible oils in the world (Gunstone, 2004). The high demand in canola oil brought about a significant increase in its production, and it is now the third most widely grown genetically modified crop after soybean and maize (Gunstone, 2004). Apart from its use as edible oil, canola oil has been converted to polyalcohols for the production of bio-based polyurethanes (Kong et al., 2012; Hong et al., 2013; Dumont et al., 2013). In addition to oil, canola contains proteins in a proportion varying from 17% to 36% (McGregor and Kimber, 1996; Newkirk and Canada, 2009). These proteins can be extracted from the meal; the by-product of oil extraction. Commercial uses of canola proteins are limited due to the presence of anti-nutritional compounds, such as glucosinolates, phenolics, and phytates (McCurdy, 1990; Manamperi et al., 2011b). Therefore, canola meal is mainly used as low value animal feed (Tan et al., 2011).

Canola proteins are characterized by a well-balanced amino acid composition, which is high in glutamic acid, aspartic acid and leucine, but low in cysteine, methionine and histidine (Klockeman et al., 1997). The similarities between canola and soy proteins suggest possible production of canola protein-based products, such as plastics, adhesives and biocomposites; which would increase the potential value of the meal. However, there are a limited number of studies related to canola protein-based films. From the available literature, it seems that very few authors dealt with blends of plasticized canola proteins and synthetic polyester by reactive extrusion (Ghorpade et al., 1995; Jang et al., 2011; Manamperi et al., 2011a). To date, there has been no report in the literature on the synthesis of canola protein-based films by solution casting.

This work addresses this knowledge gap and assesses the effects of glycerol, sodium dodecyl sulfate (SDS) and stearic acid (SA) on the properties of the canola protein-based films. The choice of SA is explained by the storage protein content of canola; 2S napin and 12S cruciferin, which accounts for 20% and 60% of total proteins, respectively (Hoeglund et al., 1992). Napin is characterized by its hydrophilic character (Aider and Barbana, 2011; Manamperi et al., 2011a). The hydrophilicity of most proteins causes water absorption in protein-based polymers which may render the resulting material unsuitable for some applications. Addition of hydrophobic lipids was found to be an effective way to reduce the hygroscopicity of these polymers (Lai et al., 1997; Lodha and Netravali, 2005a,b). Furthermore, SA acts as a plasticizer. The carboxylic group of SA can interact with the amine, imine and hydroxyl groups of canola proteins and bound to protein molecules via strong covalent bonds. On the other hand, glycerol interacts with the protein by forming

\* Corresponding author at: Department of Bioresource Engineering, McGill University, 21111 Lakeshore Rd., Ste-Anne-de-Bellevue, Québec, Canada H9X 3V9. Tel.: +1 514 398 7776; fax: +1 514 398 8387.

E-mail address: [marie-josee.dumont@mcgill.ca](mailto:marie-josee.dumont@mcgill.ca) (M.-J. Dumont).

**Table 1**  
Amino acid profile of Isolexx canola protein isolate.

Amino acids	Amino acid content (%)	
Essential	Histidine	2.45
	Isoleucine	3.03
	Leucine	7.52
	Lysine	4.74
	Methionine	1.91
	Cysteine	0.63
	Phenylalanine	4.36
	Tyrosine	3.02
	Threonine	4.13
	Valine	3.77
	Non-essential	Alanine
Arginine		6.32
Aspartic acid		8.32
Glutamic acid		17.28
Glycine		5.38
Proline		5
Serine		4.71

weak hydrogen bonds. This is called external plasticization (Lodha and Netravali, 2005b). The unreacted SA can still effectively blend with amorphous canola protein molecules, giving rise to the external plasticization (de la Caba et al., 2012). SDS is an anionic detergent which dissociates, partially unfolds and denatures proteins (Fairley et al., 1996; Huang and Sun, 2000; Mo and Sun, 2000; Zhong and Sun, 2001; Rhim et al., 2002; Katoh et al., 2004; Moosavi-Movahedi, 2005; Schmidt et al., 2005a,b; Sessa et al., 2006). The modification of the protein's structure by denaturation allows the three-dimensional structure of the protein to completely or partially unfold without breaking the peptide bonds and improve the performance of these polymers (De Graaf, 2000).

We found that there are significant differences in the mechanical, thermal, structural and water absorption properties of the films depending on the plasticizers and additives used for their synthesis.

## 2. Materials and methods

### 2.1. Material

Canola protein isolates (CPI) (Isolexx Canola Protein, 93% protein content on dry basis) were obtained from BioExx (Toronto, ON, Canada). The amino acid profile of CPI is shown in Table 1. NaOH was obtained from EMD (Damstadt, Germany), glycerol was purchased from J.T.Baker (Phillipsburg, NJ, US), SA was purchased from SAFC (St.-Louis, MO, US) and SDS was purchased from Fisher Scientific (Fair Law, NJ, US).

### 2.2. Film formation

Glycerol was added to 10 g  $\pm$  0.1 mg of CPI at different ratios (25, 30, 35, 40 wt%). The film-forming solutions were prepared by slowly dissolving the mixture of CPI and glycerol into 200 g of continuously stirred distilled water. The pH of the solutions was adjusted to 11.0  $\pm$  0.3 using a 1 M NaOH solution. The solutions were heated in a water bath while being stirred at 70 °C for 20 min. The solutions were cooled to room temperature (23  $\pm$  2 °C) in a cold water bath. The solutions were degassed by applying a vacuum, and casted onto silicone plates (22 cm  $\times$  22 cm). The film thickness was controlled by casting the same amount (210 mL) of solution per plate. The plates were placed on a leveled surface in a ventilated climate chamber (E15 CONVIRON, Winnipeg, MB, Canada) for 48 h at 25 °C and 30% RH to allow for solvent evaporation. The dried films were peeled off from the pans. Similarly, films were prepared by adding SDS or SA at 5, 10 and 15 wt% to the CPI and the glycerol prior to pH adjustment. The compositions of the films are given in Table 2.

**Table 2**  
Composition of CPI-based films.

Films	GLY (wt%)	SDS (wt%)	SA (wt%)	GLY (g)	SDS (g)	SA (g)
CPI 25GLY	25	0	0	3.10	0.00	0.00
CPI 30GLY	30	0	0	3.98	0.00	0.00
CPI 35GLY	35	0	0	5.01	0.00	0.00
CPI 40GLY	40	0	0	6.20	0.00	0.00
CPI 25GLY 5SDS	25	5	0	3.32	0.66	0.00
CPI 25GLY 10SDS	25	10	0	3.58	1.43	0.00
CPI 25GLY 15SDS	25	15	0	3.88	2.33	0.00
CPI 30GLY 5SDS	30	5	0	4.29	0.72	0.00
CPI 30GLY 10SDS	30	10	0	4.65	1.55	0.00
CPI 30GLY 15SDS	30	15	0	5.07	2.54	0.00
CPI 35GLY 5SDS	35	5	0	5.43	0.78	0.00
CPI 35GLY 10SDS	35	10	0	5.92	1.69	0.00
CPI 35GLY 15SDS	35	15	0	6.51	2.79	0.00
CPI 40GLY 5SDS	40	5	0	6.76	0.85	0.00
CPI 40GLY 10SDS	40	10	0	7.44	1.86	0.00
CPI 40GLY 15SDS	40	15	0	8.27	3.10	0.00
CPI 25GLY 5SA	25	0	5	3.32	0.00	0.66
CPI 25GLY 10SA	25	0	10	3.58	0.00	1.43
CPI 25GLY 15SA	25	0	15	3.88	0.00	2.33
CPI 30GLY 5SA	30	0	5	4.29	0.00	0.72
CPI 30GLY 10SA	30	0	10	4.65	0.00	1.55
CPI 30GLY 15SA	30	0	15	5.07	0.00	2.54
CPI 35GLY 5SA	35	0	5	5.43	0.00	0.78
CPI 35GLY 10SA	35	0	10	5.92	0.00	1.69
CPI 35GLY 15SA	35	0	15	6.51	0.00	2.79
CPI 40GLY 5SA	40	0	5	6.76	0.00	0.85
CPI 40GLY 10SA	40	0	10	7.44	0.00	1.86
CPI 40GLY 15SA	40	0	15	8.27	0.00	3.10

### 2.3. Film characterization

#### 2.3.1. Fourier transformed infrared (FT-IR)

The composition of CPI powder, CPI films (30 GLY, 30 GLY-SA and 30 GLY-SDS) and additives were measured in triplicate on a Nicolet iS5 FT-IR spectrometer (Thermo, Madison, WI, USA). The spectra were recorded at 32 scans and 4 cm<sup>-1</sup> resolution in the 4000–400 cm<sup>-1</sup> range. The spectra were analyzed using the OMNIC software package (version 8.2, Thermo Nicolet Corp.).

#### 2.3.2. X-ray diffraction

The effects of additives (SDS and SA) on the structure of the CPI-based films were investigated by X-ray diffraction. The X-ray patterns were obtained by using an AXS X-ray diffractometer (Bruker, Madison, WI) with Cu-K $\alpha$  radiation ( $\lambda$  = 0.1542 nm) at 45 kV and 40 mA. Films were scanned at room temperature in the range of diffraction angle of  $2\theta$  = 3–70°.

#### 2.3.3. Thermogravimetric analysis

The thermal stability of the films was determined using a thermogravimetric analyzer (TGA) (Q500, TA Instrument, Inc., New Castle, DE). The thermogravimetric analyses were carried out under a stream of nitrogen at a flow rate of 60 mL/min. Each sample weighed between 5 to 10 mg and was heated from room temperature to 600 °C at a constant rate of 20 °C/min. The CPI and films (30 GLY, 30 GLY-10 SA and 30 GLY-10 SDS) were tested in duplicate.

#### 2.3.4. Differential scanning calorimetry (DSC)

A 10 mg quantity of CPI powder and films (30 GLY, 30 GLY-SA and 30 GLY SDS) were compressed in hermetic aluminum pans and scanned in duplicate using a DSC (Q100, TA Instruments, Inc., New Castle, DE) under a stream of nitrogen (50 mL/min). Samples were initially heated from 0 °C to 200 °C at a rate of 10 °C/min to remove previous thermal history. Samples were then cooled to 0 °C at a rate of 10 °C/min. The samples were reheated from 0 °C to 250 °C at a rate of 10 °C/min.



### 2.3.5. Dynamic mechanical analysis

The viscoelastic behavior of CPI-based film were examined by using a dynamic mechanical analyzer (DMA) (D 100, TA Instrument, Inc., New Castle, DE), which is equipped with a nitrogen cooling system. Films that were 17.5 mm long, 12 mm wide and 0.6 mm thick were analyzed in a single cantilever bending mode at a frequency of 1 Hz and a fixed oscillation displacement of 15  $\mu\text{m}$ . The specimens were heated at a constant rate of 1  $^{\circ}\text{C}/\text{min}$  between  $-100^{\circ}\text{C}$  and  $40^{\circ}\text{C}$ . The results were analyzed using the Universal Analysis 2000 software (TA Instruments, Inc., New Castle, DE).

### 2.3.6. Mechanical properties

The mechanical properties (elongation at break and tensile strength) were tested using an Instron Universal Testing Machine (model 4500, Instron Corporation, Canton, MA) according to ASTM standard 638-10. Film specimens were cut into dog bone shapes, 3.18 mm wide and 9.53 mm long. The thicknesses of the specimen were measured at three random places on the narrow section of the film using a micrometer (030025, Marathon watch company Ltd., Richmond Hill, Ontario). The average thickness was calculated and recorded. Film specimens were conditioned at  $23^{\circ}\text{C}$  and 50% relative humidity for 40 h prior to the tests. The tests were conducted at a cross head rate of 10 mm/min. Seven replicates were tested for each blend and the average values and corresponding standard deviations are presented.

### 2.3.7. Water absorption

The water absorptivity of the films (30 GLY, 30 GLY-SA and 30 GLY-SDS) were determined by the ASTM standard method D570-98. Films were cut into rectangular specimens measuring 76.2 mm in length and 25.4 mm in width, and were conditioned in an isothermal oven (Model 1327F, VWR, Cornelius, OR) at  $50^{\circ}\text{C}$  for 24 h. The specimens were cooled to room temperature in a desiccator and weighed. The preconditioned specimens were immersed in distilled water at  $25^{\circ}\text{C}$  for 2 h, and weighed after removing any surface water. The wet specimens were conditioned again at  $50^{\circ}\text{C}$  for 24 h and weighed to calculate the loss of soluble matter. The total water absorption was calculated as the sum of water absorbed (Eq. (1)) and the soluble matter loss (Eq. (2)). Samples were tested in triplicate and average values and corresponding standard deviations were determined.

$$\text{Increase in weight (\%)} = \frac{\text{wet weight} - \text{conditioned weight}}{\text{conditioned weight}} \times 100 \quad (1)$$

$$\begin{aligned} \text{Soluble matter lost (\%)} \\ = \frac{\text{conditioned weight} - \text{reconditioned weight}}{\text{conditioned weight}} \times 100 \end{aligned} \quad (2)$$

### 2.3.8. Morphological property

The morphological properties of CPI films were examined using a JEOL JSM-7600 TFE scanning electronic microscope (SEM) (JEOL Ltd., Tokyo, Japan). The examinations were carried out at a magnification of  $5000\times$  at an accelerated voltage of 2 kV. The surface of the films was coated with gold using a Polaron SC-502 sputter coater (Fison, Ashford, UK) under an argon atmosphere. The coating process was conducted at a plasma current of 5 mA for 30 s so as to have a coating thickness of less than 10 nm.

The microstructure of the films was also observed using a polarized light microscope, PLM, (Leica DMRX, Leica Microsystem, Wetzlar, Germany) fitted with a Hamamatsu (C474295) digital camera for image capture. The observations were made at a magnification of  $10\times$  at room temperature.

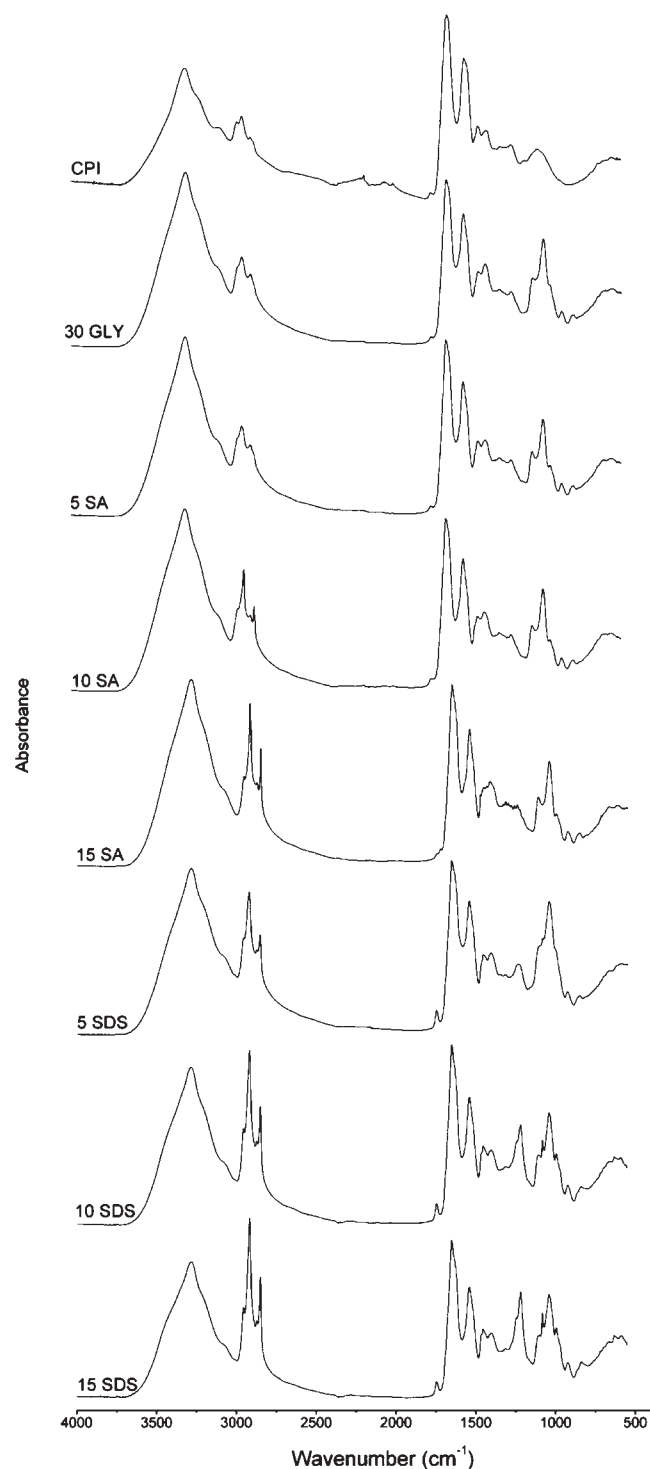


Fig. 1. FT-IR spectra for CPI powder, CPI films and modified CPI films.

## 3. Results and discussion

### 3.1. FT-IR

The FT-IR spectra of CPI powder and CPI-based films are shown in Fig. 1. For the powder; amide I band, amide II band and C–H band were observed at  $1643\text{ cm}^{-1}$ ,  $1537\text{ cm}^{-1}$ , and  $1447\text{ cm}^{-1}$ , respectively. The broad absorbance ranging from wavenumbers 3200 to  $3600\text{ cm}^{-1}$  was attributable to the free and bound O–H and



N–H groups. The absorption band at wavenumber  $2927\text{ cm}^{-1}$  was attributed to the stretching of C–H in  $\text{CH}_2$  and  $\text{CH}_3$  groups.

Compared with the CPI, there were no significant changes in the CPI-based films plasticized with 30 wt% of glycerol (Fig. 1), except a clear increase in the intensity of the absorption band associated with the free O–H groups of glycerol. Moreover, the amide I band changed to higher wavenumber from  $1644\text{ cm}^{-1}$  to  $1650\text{ cm}^{-1}$ , suggesting that a rearrangement of the protein structure had occurred, such as the dissociation of peptides into subunits (Schmidt et al., 2005b). The change in the structure of the proteins could have resulted from the high temperature and the high alkalinity of the film forming solution at the film processing stage. Furthermore, the hydroxyl and amino groups of the CPI may have been able to form inter- or intra-hydrogen bonding with the carbon–oxygen bond on amino acid due to moisture (Schmidt et al., 2005b). As the CPI was processed into films, an acetylenic bond deformation at wavenumber  $2160\text{ cm}^{-1}$  was observed (Solomons et al., 2011).

The films containing SA showed a significant increase in the absorbance band at wavenumbers  $2915\text{ cm}^{-1}$  and  $2849\text{ cm}^{-1}$  (Fig. 1). The increments agreed with the spectra of SA (data not shown) where two clear peaks were observed at wavenumbers  $2913\text{ cm}^{-1}$  and  $2847\text{ cm}^{-1}$ . These wavenumbers are within the range of the O–H stretching in carboxylic acid (between  $2500\text{ cm}^{-1}$  and  $3000\text{ cm}^{-1}$ ) (Solomons et al., 2011). Theoretically, the formation of new linkages induced by SA should result in changes between wavenumbers  $1600$  and  $800\text{ cm}^{-1}$  (Lodha and Netravali, 2005b). However, no significant variation was observed in this region. Under alkaline conditions, the carboxylic acid group of SA could hypothetically react with CPI to form new amide linkages and could also ionize into carboxylate ( $-\text{CO}_2^-$ ) with an expected peak absorbance at wavelength  $1600\text{ cm}^{-1}$ . But this peak was probably masked by that for the amide linkages (Lodha and Netravali, 2005b).

From the spectra of films modified with SDS, increases in absorbance at wavenumbers  $2850\text{--}2920\text{ cm}^{-1}$  and at  $1745\text{ cm}^{-1}$  were observed (Fig. 1) as compared to films plasticized with glycerol. This indicated the formation of free C–H and ester bonds respectively. The decrease in the absorbance of amide I and II bands could have resulted from the unfolding or partially unfolding of the protein structure caused by SDS. The free peptide bonds probably formed new linkages with glycerol or with other peptides inside of the unfolded protein structure and consequently, the absorbance of the band in the  $2850\text{--}2920\text{ cm}^{-1}$  range decreased. The increase in the band absorbance for free C–H stretching was attributed to the participation of SDS, which has spectra with two clear peaks at wavenumbers,  $2849\text{ cm}^{-1}$  and  $2916\text{ cm}^{-1}$  (data not shown). The formation of a peak at wavenumber number  $1220\text{ cm}^{-1}$  was induced by the sulfate group of SDS. As the concentration of SDS increased, the intensity of the peak increased.

### 3.2. X-ray diffraction (XRD)

The X-ray diffraction patterns of CPI-based films, films modified by 10 wt% of SDS and film co-plasticized by 10 wt% of SA are shown in Fig. 2. No distinct peak was observed in the XRD pattern of 30 GLY films, indicating the CPI films plasticized by glycerol alone is amorphous. However, the XRD pattern of 30 GLY 10 SDS films showed several distinct peaks at  $2\theta$  values of  $4.5^\circ$ ,  $6.8^\circ$ , and  $11.1^\circ$ , which have been reported positions of characteristic peaks for pure SDS (Bittencourt, 1988). When compared with the XRD patterns of 30 GLY films and 30 GLY 10 SDS films, more distinct peaks were observed in the patterns for films co-plasticized with 10 wt% of SA. The peak positions for the 30 GLY 10 SA films were different from those in the XRD pattern of the pure SA crystals (Lodha and Netravali, 2005a), suggesting the formation of new

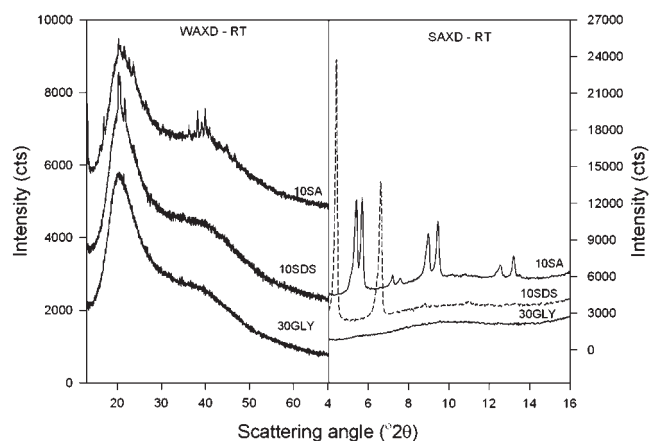


Fig. 2. XRD analysis of CPI films: left, wide angle XRD pattern; right, small angle XRD pattern.

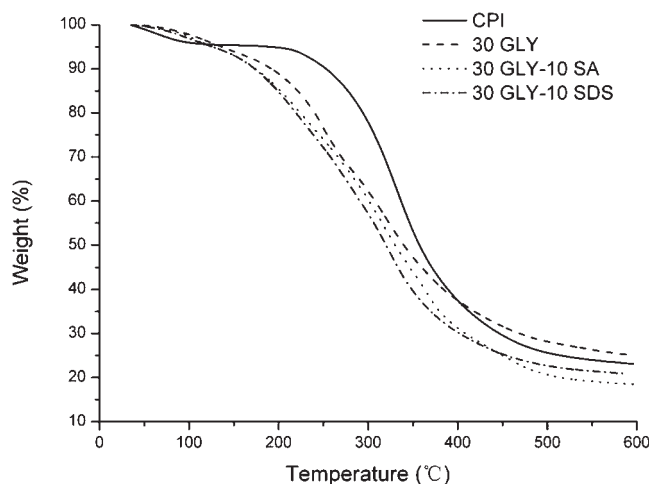


Fig. 3. Thermogravimetric analysis of CPI and canola protein-based films.

ordered structures from the reaction between SA, glycerol, and CPI during film processing.

### 3.3. Thermal properties

#### 3.3.1. Thermo-gravimetric analysis

The TGA scans of CPI and CPI-based films plasticized with 30 wt% glycerol are shown in Fig. 3. A major decrease in thermal stability was observed when CPI is plasticized with glycerol and additives (Table 3). This was attributed to the high vapour pressure of glycerol, which starts degrading at about  $104^\circ\text{C}$  (TGA data not shown). The changes in the protein structure due to plasticization could also lead to the observed changes in thermal stability. Fig. 3b shows that the degradation of CPI has two major stages of mass loss. The first stage, observed up to  $120^\circ\text{C}$ , is associated with the moisture content. The second stage is related to the denaturation of the protein

Table 3

Thermogravimetric parameters for the thermal degradation of CPI and CPI-based films.

Sample	5% Degradation temperature ( $^\circ\text{C}$ )	10% Degradation temperature ( $^\circ\text{C}$ )	Ash content (%)
CPI	179	249	23
30 GLY	126	181	24
30 GLY-10 SA	124	174	19
30 GLY-10 SDS	124	173	21

and this stage can be further divided into two degradation phases based on the derivative TGA curve (Fig. 4). The existence of two degradation phases was attributed to the two fractions of canola proteins. The first stage between temperature 100 °C and 150 °C and the second stage from 220 °C to 250 °C are considered as the denaturation of napin and cruciferin respectively, which agreed with the transition temperature exhibited in the DSC scans (Fig. 5).

### 3.3.2. Differential scanning calorimetry

The DSC scans of pure SA, SDS, CPI powder, and selected CPI-based films are presented in Fig. 5. The powder showed two endothermic transitions. The first was observed in the first heating cycle at 111 °C with an enthalpy of 154 J/g. The second transition was observed in the second heating cycle at 217 °C with an enthalpy of 6 J/g. The two endothermic transitions correspond to the denaturation of 2S napin and 12S cruciferin, respectively. Compared to the powder, the DSC scans of 30 GLY showed shifts in the denaturation temperatures (133 °C and 237 °C) and a decrease in the napin and cruciferin denaturation enthalpies (111 J/g and 1 J/g, respectively). The decrease of enthalpies indicated the reduction in the

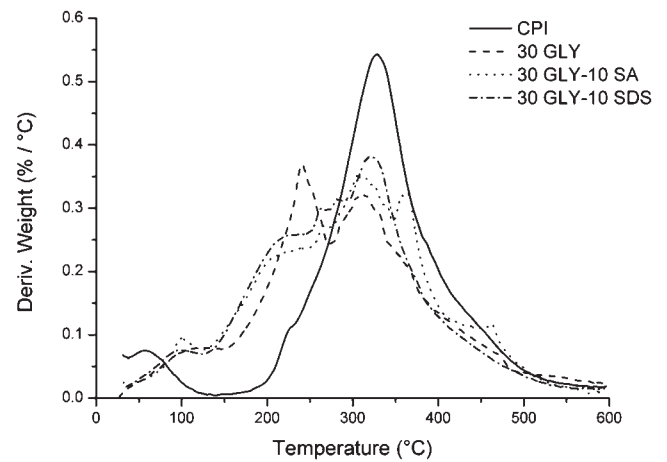


Fig. 4. Derivative TGA curve for CPI and CPI-based films.

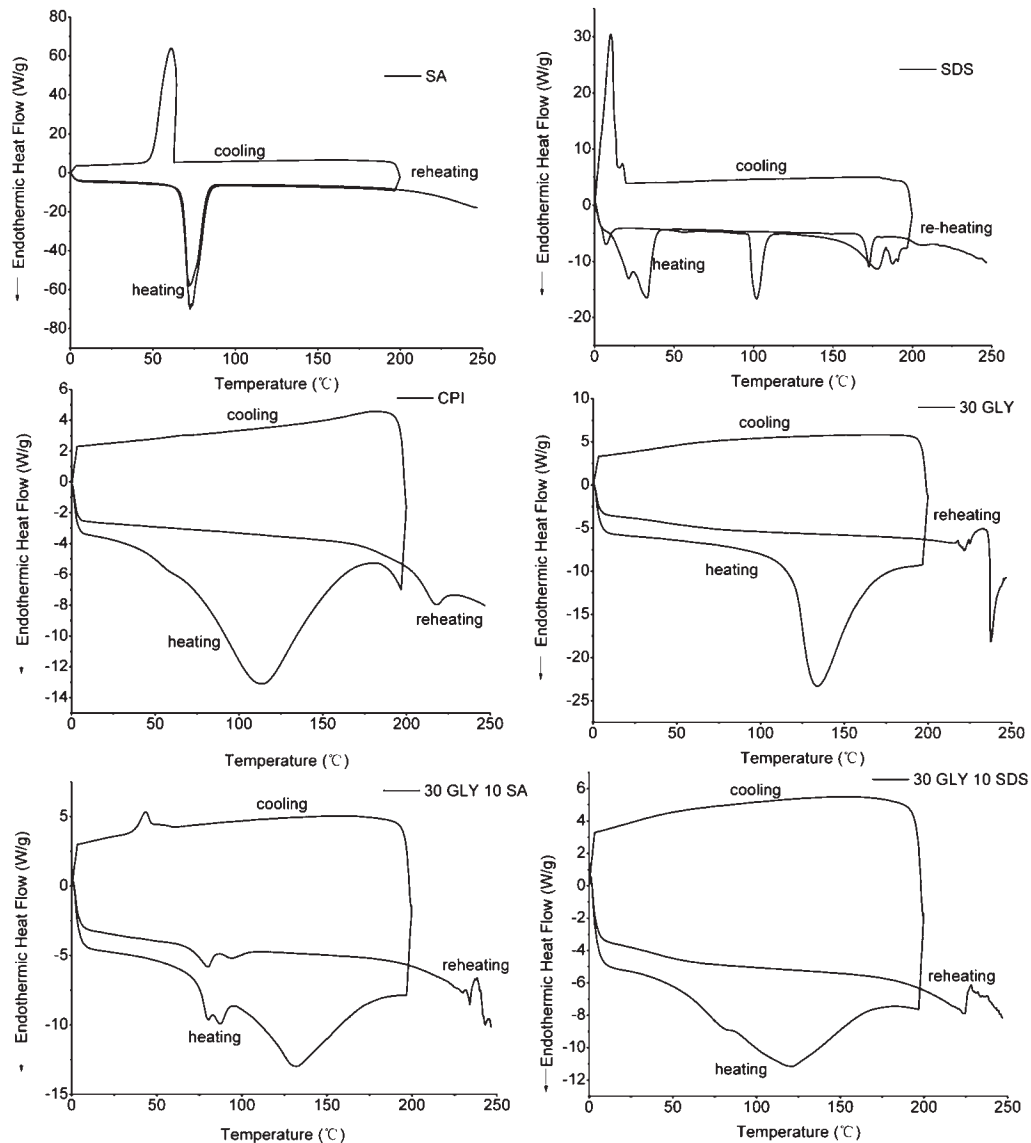


Fig. 5. DSC thermograms of CPI, CPI films and modified CPI films.

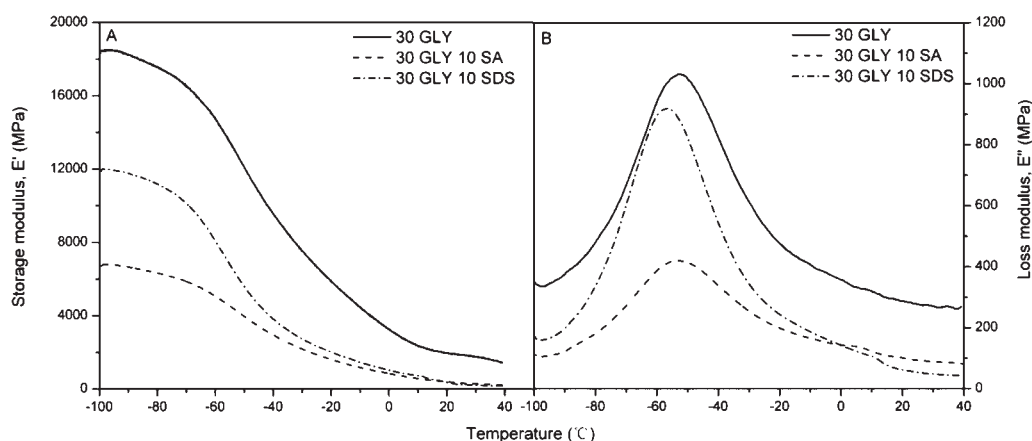


Fig. 6. Dynamic mechanical properties of CPI based films, (A) storage modulus  $E'$  and (B) loss modulus  $E''$ .

thermal stability of CPI-based films, which may have resulted from the plasticization with glycerol.

The DSC scans of CPI-based films co-plasticized with SA (30 GLY SA) presented two denaturation transitions at 132 °C and 234 °C. It is worth mentioning that an exothermic transition observed in the cooling stage of 30 GLY SA is similar to the transition found in the thermogram of pure SA for the same cycle stage. Comparing the enthalpy of the exothermic transition of the 10 SA films to the exothermic transition enthalpy of pure SA, only 1.7% of SA was present in its crystalline form, meaning that the majority of SA (98%) may have reacted with glycerol or dispersed within the polymeric matrices. This agrees with the XRD analysis (Fig. 2). Similar DSC analyses were also reported for soy protein-based resin films co-plasticized with SA (Lodha and Netravali, 2005b).

Decreases in the denaturation temperature as well as the protein denaturation enthalpies for both protein fractions were observed in the DSC scans of CPI-based films with the addition of SDS (30 Gly SDS). Compared to 30 GLY, the denaturation temperature of napin and cruciferin decreased from 133 °C and 237 °C to 120 °C and 224 °C, respectively. The reduction in these aforementioned values indicated a decrease in thermal stability through addition of SDS, which was attributed to the high degree of protein denaturation and the formation of disordered protein structures.

### 3.4. Dynamic mechanical properties

Storage and loss moduli of selected CPI-based films are presented in Fig. 6A and B respectively. Glass transition temperature ( $T_g$ ) is an important characteristic which can be derived from DMA testing. The dissipation energy reaches its maximum at  $T_g$ , which results in a maximum in the loss modulus curve (Mo and Sun, 2000). Therefore, from Fig. 3B, the  $T_g$  for 30 GLY, 10 SA and 10 SDS films were observed at temperatures of -52.7 °C, -53.2 °C and -56.8 °C, respectively. The lowest  $T_g$ 's were observed for films containing SDS, which may result from an increase in the protein chain flexibility induced by protein denaturation (Mo and Sun, 2000). The water absorption capacity of the films containing SDS was higher than films plasticized with glycerol and co-plasticized with SA (Fig. 8). The high moisture content of these films may have also led to the decrease in  $T_g$  since water acts as a plasticizer.

### 3.5. Mechanical properties

The mechanical properties of CPI-based films are shown in Table 4. As a general trend for films plasticized with glycerol (Group 1), the elongation of the films at break increased and the tensile strength decreased as the glycerol content increased. This tendency

Table 4

Mechanical properties of CPI-based films processed by casting.

Group	Film	Elongation at break (%)	Tensile strength (MPa)
1	25 GLY	41.9 ± 12.3	1.7 ± 0.3
	30 GLY	163 ± 6.4	1.4 ± 0.2
	35 GLY	207.1 ± 28.2	1.1 ± 0.3
	40 GLY	287.4 ± 8.3	0.7 ± 0.1
2	25 GLY 5 SA	28 ± 2.3	1.5 ± 0.1
	25 GLY 10 SA	12.3 ± 3.1	2.2 ± 0.5
	25 GLY 15 SA	8.8 ± 2.7	2.1 ± 0.1
	30 GLY 5 SA	43.8 ± 7.3	0.8 ± 0.2
	30 GLY 10 SA	25.5 ± 1.6	1.2 ± 0.2
	30 GLY 15 SA	23.3 ± 0.9	0.8 ± 0.1
	35 GLY 5 SA	88.4 ± 13.1	0.5 ± 0.1
	35 GLY 10 SA	36.9 ± 3.5	0.5 ± 0.1
	35 GLY 15 SA	26.1 ± 3.7	0.6 ± 0.2
	40 GLY 5 SA	128.9 ± 20.5	0.4 ± 0.1
3	40 GLY 10 SA	50.3 ± 8.9	0.2 ± 0
	40 GLY 15 SA	28.2 ± 2.1	0.2 ± 0.1
	25 GLY 5 SDS	115.4 ± 48.1	2.3 ± 0.3
	25 GLY 10 SDS	69.2 ± 14.3	2.6 ± 0.2
	25 GLY 15 SDS	24.6 ± 3.2	3.8 ± 0.2
	30 GLY 5 SDS	199.4 ± 32.8	2.1 ± 0.3
	30 GLY 10 SDS	244.3 ± 62.7	2 ± 0.2
	30 GLY 15 SDS	196.5 ± 26.2	1.8 ± 0.1
	35 GLY 5 SDS	450.2 ± 39.5	1 ± 0.1
	35 GLY 10 SDS	488.1 ± 46.1	1.1 ± 0.1
35 GLY 15 SDS	364.2 ± 52.5	1.2 ± 0.1	
40 GLY 5 SDS	551.4 ± 48.9	1.2 ± 0.1	
40 GLY 10 SDS	542.8 ± 63.3	1.2 ± 0.1	
40 GLY 15 SDS	446 ± 29.3	0.9 ± 0	

is common for protein matrices plasticized with glycerol (Lodha and Netravali, 2005b; Guerrero et al., 2010). The addition of glycerol could have reduced the interactions between protein chains and even break protein's intermolecular linkage, which stabilizes the primitive structure of protein through the interactions between the hydroxyl groups of glycerol and amino acids (Guerrero et al., 2010). Therefore, the mobility of the protein chains increased, resulting in an increase in elongation at break and a decrease in tensile strength. The standard deviation related to the elongation at break of films plasticized with 25 wt% glycerol is higher than the films plasticized with a higher ratio of glycerol due to incomplete plasticization.

The addition of SA as a co-plasticizer (Group 2) showed that both elongation at break and tensile strength decreased for the same glycerol content. The addition of SA had a drastic effect on the elongation at break values when compared to the values found for Group 1. For example, the addition of 15 wt% of SA to matrices plasticized with 40 wt% glycerol decreased elongation at break by 90%. These results are in agreement with the studies by Lodha and Cho

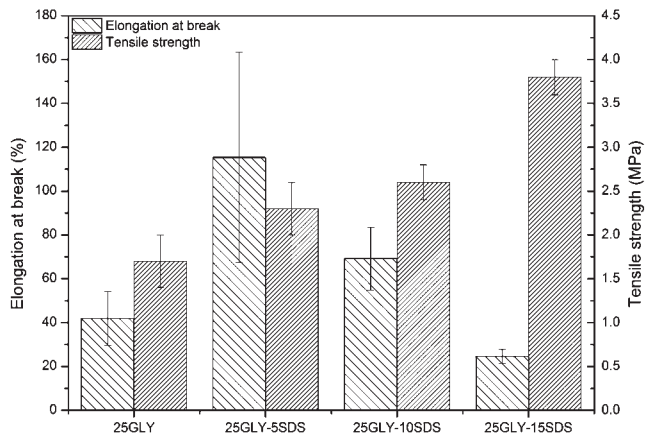


Fig. 7. Mechanical properties of films formulated with 25 wt% of GLY and various ratios of SDS.

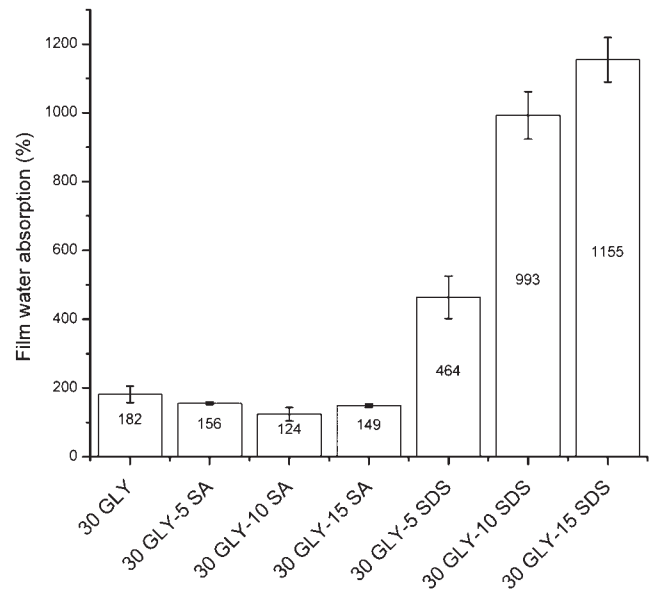


Fig. 8. Water absorption of CPI-based films.

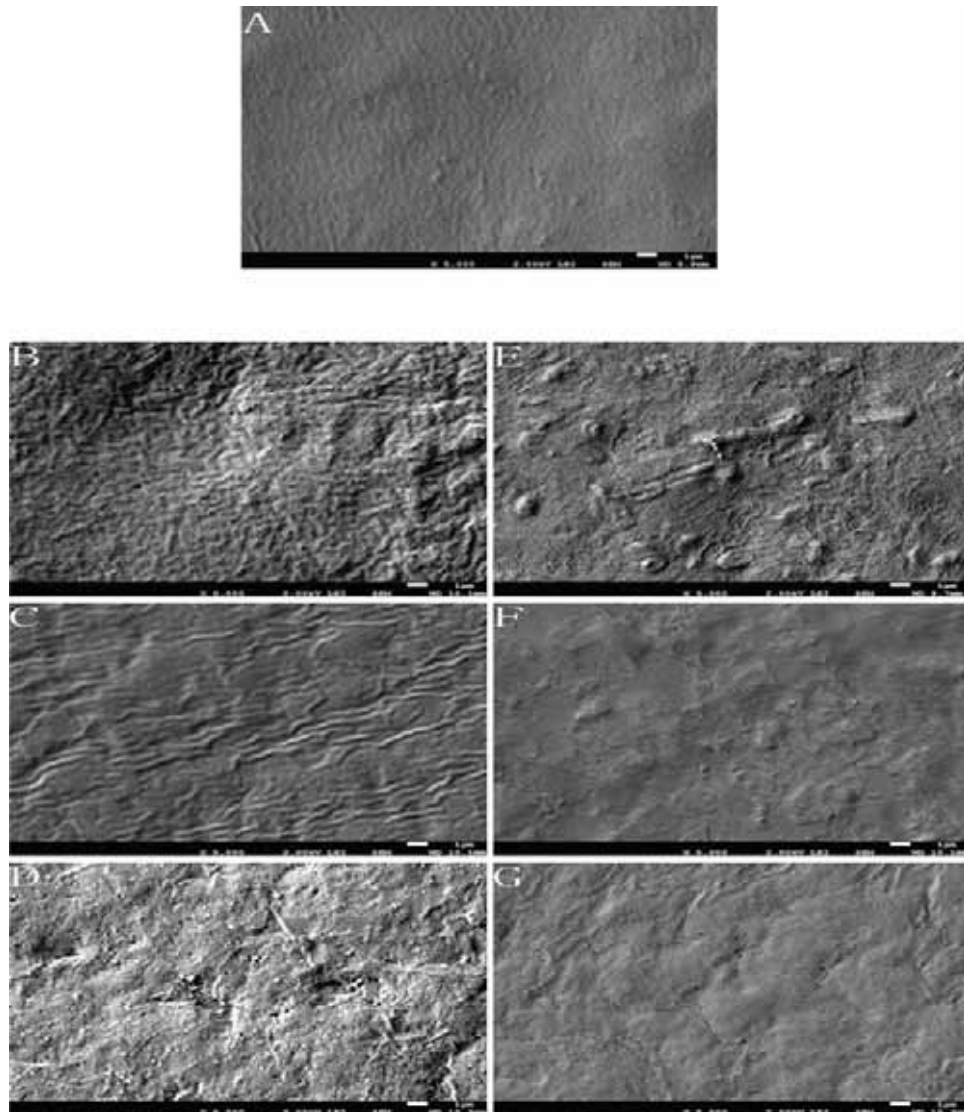


Fig. 9. SEM images of CPI films and modified CPI films: (A) 30 GLY, (B) 30 GLY 5 SA, (C) 30 GLY 10 SA, (D) 30 GLY 15 SA, (E) 30 GLY 5 SDS, (F) 30 GLY 10 SDS and (G) 30 GLY 15 SDS.



(Cho et al., 2002; Lodha and Netravali, 2005b), in which the elongation at break and the tensile strength of soy protein-based films and whey protein-based films decreased with increasing concentrations of SA. SA is a hydrophobic fatty acid which may decrease the amount of water absorbed during conditioning. Water is a powerful plasticizer and a decrease in water content would have resulted in an increase in brittleness.

The addition of SDS increased the elongation at break values of the films (Group 3) by as much as 175% for films formulated with 5 wt% SDS and 25 wt% glycerol when compared to films plasticized with 25 wt% glycerol as presented in Fig. 7. The tensile strength also increased for all films except for the case of 35 wt% glycerol. There was no SDS concentration-dependent response observed for changes in the tensile strength. However, as the concentration of SDS increased from 10 wt% to 15 wt% for each level of glycerol, the elongation of all films decreased. SDS is an anionic detergent or surfactant and acts as a powerful denaturant and dissociating agent for proteins (Huang and Sun, 2000; Rhim et al., 2002). In its original state, the protein's structure is stabilized by intermolecular hydrophobic interactions (e.g. disulphide bonds) and hydrogen bonds. Dissociation of proteins by SDS occurs when the non-polar portion of SDS reacts with the hydrophobic amino acid residues, thereby disrupting hydrophobic interactions within the protein. In addition to the hydrophobic interferences, the anionic portion of the SDS disrupts electrostatic interactions within the protein which changes the secondary, tertiary, and quaternary structures and conformation of the protein, which partially unfolds the protein and increases its flexibility (Manamperi and Pryor, 2012). Since the pH of the film forming solution was higher than the isoelectric point of the canola proteins, the protein molecules were negatively charged. The electrostatic repulsion imposed by SDS and the negative charge of the proteins were the driving forces for the change in configuration. This change in configuration brought an improved contact area between the proteins and the plasticizer, which resulted in an increase in mechanical properties (Rhim et al., 2002). However, at higher SDS concentration (15 wt%), the excessive SDS may have disrupted the protein matrix and reduce the interactions between the peptide chains and consequently decreased the elongation at break. Similarly, Mo et al. also reported a decrease in mechanical properties of soy protein films at higher SDS concentration (Mo and Sun, 2000).

### 3.6. Water absorption

The water absorption capacities of the CPI-based films are presented in Fig. 8. Compared to films plasticized with glycerol, there is a small decrease in the amount of water absorbed when SA is added as a co-plasticizer. The increase in water resistance was attributed to the hydrophobicity of SA. A significant increase in water absorption occurred when the films were modified with SDS. This observation is consistent with the literature, which reported that SDS modification improved the water absorption capacities of soy protein based films and of films made from blends of canola protein and polyesters (Mo and Sun, 2000; Zhong and Sun, 2001; Manamperi and Pryor, 2012). The high water absorption of SDS modified CPI films probably resulted from the denaturation of the hydrophobic protein proportion – cruciferin. Once the cruciferin was denatured, the water resistance of the canola protein decreased, and thus the resulting film would absorb more water (Zhong and Sun, 2001).

### 3.7. Morphological properties

The results of investigation of selected CPI-based films by SEM are presented in Fig. 9. From this figure, the 30 GLY film is characterized by a homogeneous surface. With the addition of SA, more

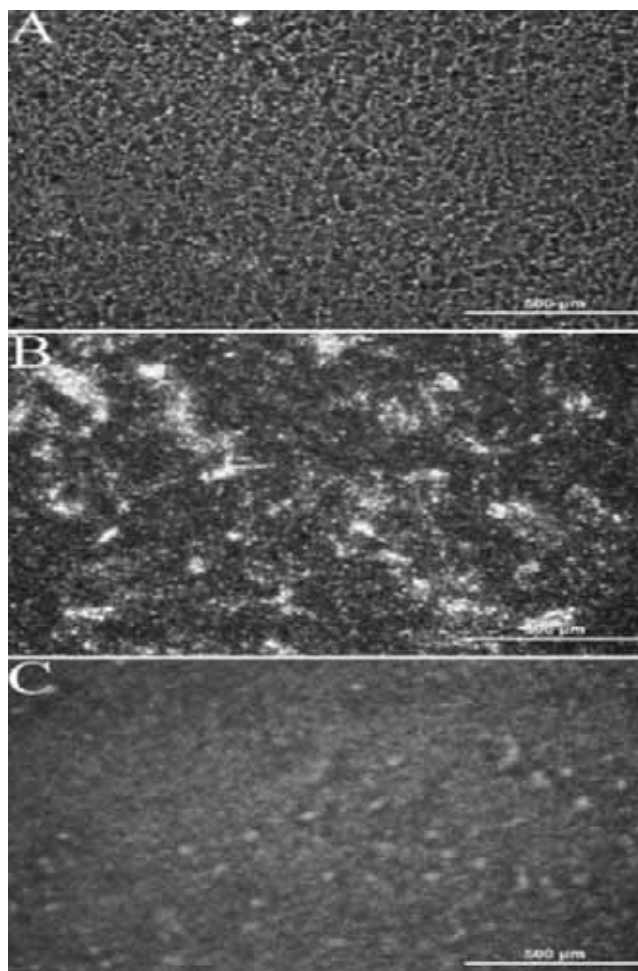


Fig. 10. Polarized light microscopy of CPI films under 10× magnification: (A) 30 GLY, (B) 30 GLY-10 SA and (C) 30 GLY-10SDS.

grooves appeared on the surfaces of the films as compared to 30 GLY films. After increasing the concentration of SA from 5 to 10 wt%, the surface of the films was smoother but long creases were present. As the SA concentration was increased to 15 wt%, the film presented a rougher surface with numerous fissures. These results agreed with the results obtained for tests of mechanical properties (Table 4), for which the films with the highest tensile strength and elongation at break was those with 10 wt% SA, as compared to films with other concentrations of SA added.

The surface morphology of 30 GLY 10 SDS is different from other films as cluster-shaped structures were observed. As the concentration of SDS increased from 5 to 10 wt%, a decrease in the concentration of clusters occurred. With a SDS concentration of 15 wt%, the surface of the films became smoother. This might be due to a greater degree of protein denaturation.

PLM analyses of selected films are shown in Fig. 10. Considerable amounts of crystal-like structures were observed in films after SA addition, which is different from the 30 GLY and 30 GLY 10 SDS. These results agreed with the XRD analysis (Fig. 2). Fewer crystal-like structures were observed for films after SDS addition, which was probably the result of the high degree of protein denaturation.

## 4. Conclusion

The results of this research showed that canola proteins can be utilized as feedstock for the manufacture of biodegradable polymers. This study reported great differences in the mechanical,



structural, thermal and water absorption properties of the polymeric matrices depending on the nature of the plasticizer and additives being used. According to the results, films modified by 10 wt% of SDS at various glycerol concentrations afforded the highest elongation at break values, indicating that SDS is an effective additive to improve the mechanical properties of canola protein-based films. It was a surprise that the water absorption capacity of some of these matrices could reach as high as 1115%, suggesting that these proteins could be used in the manufacture of superabsorbent.

### Acknowledgements

This research was financially supported by the Start-Up Fund of McGill University. We gratefully acknowledge the permission of using the laboratory equipment of Dr. Suresh S. Narine (Trent University, Ontario), Dr. Valérie Orsat (McGill University, Québec) and Dr. Michael Ngadi (McGill University, Québec). We also thank Mr. Yvan Garipey for technical support and Dr. Darwin Lyew for help in editing.

### References

- Aider, M., Barbana, C., 2011. Canola proteins: composition, extraction, functional properties, bioactivity, applications as a food ingredient and allergenicity – a practical and critical review. *Trends in Food Science and Technology* 22, 21–39.
- Bittencourt, D.R.S., 1988. X-ray powder diffraction data for sodium octyl sulfate, sodium decyl sulfate and sodium dodecyl sulfate. *Powder Diffraction* 3, 244–245.
- Cho, S.Y., Park, J.W., Rhee, C., 2002. Properties of laminated films from whey powder and sodium caseinate mixtures and zein layers. *Lebensmittel-Wissenschaft & Technologie-Food Science and Technology* 35, 135–139.
- De Graaf, L.A., 2000. Denaturation of proteins from a non-food perspective. *Journal of Biotechnology* 79, 299–306.
- de la Caba, K., Pena, C., Ciannamea, E.M., Stefani, P.M., Mondragon, I., Ruseckaite, R.A., 2012. Characterization of soybean protein concentrate-stearic acid/palmitic acid blend edible films. *Journal of Applied Polymer Science* 124, 1796–1807.
- Dumont, M.-J., Kharraz, E., Qi, H., 2013. Production of polyols and mono-ols from 10 North-American vegetable oils by ozonolysis and hydrogenation: a characterization study. *Industrial Crops and Products* 49, 830–836.
- Fairley, P., Monahan, F.J., German, J.B., Krochta, J.M., 1996. Mechanical properties and water vapor permeability of edible films from whey protein isolate and sodium dodecyl sulfate. *Journal of Agricultural and Food Chemistry* 44, 438–443.
- Ghorpade, V.M., Li, H., Gennadios, A., Hanna, M.A., 1995. Chemically modified soy protein films. *Transactions of the ASAE* 38, 1805–1808.
- Guerrero, P., Retegi, A., Gabilondo, N., de la Caba, K., 2010. Mechanical and thermal properties of soy protein films processed by casting and compression. *Journal of Food Engineering* 100, 145–151.
- Gunstone, F.D., 2004. *Rapeseed and Canola Oil: Production, Processing, Properties and Uses*. Blackwell Pub./CRC Press, Oxford, EN/Boca Raton, FL.
- Hoeglund, A.S., Roedin, J., Larsson, E., Rask, L., 1992. Distribution of napin and cruciferin in developing rape seed embryos. *Plant Physiology* 98, 509–515.
- Hong, J., Shah, B.K., Petrovic, Z.S., 2013. Vegetable oil cast resins via click chemistry: effects of cross-linkers. *European Journal of Lipid Science and Technology* 115, 55–60.
- Huang, W., Sun, X., 2000. Adhesive properties of soy proteins modified by sodium dodecyl sulfate and sodium dodecylbenzene sulfonate. *Journal of the American Oil Chemists' Society* 77, 705–708.
- Jang, S.-A., Lim, G.-O., Song, K.B., 2011. Preparation and mechanical properties of edible rapeseed protein films. *Journal of Food Science* 76, C218–C223.
- Katoh, K., Shibayama, M., Tanabe, T., Yamauchi, K., 2004. Preparation and physicochemical properties of compression-molded keratin films. *Biomaterials* 25, 2265–2272.
- Klockeman, D.M., Toledo, R., Sims, K.A., 1997. Isolation and characterization of defatted canola meal protein. *Journal of Agricultural and Food Chemistry* 45, 3867–3870.
- Kong, X.H., Liu, G.G., Curtis, J.M., 2012. Novel polyurethane produced from canola oil based poly(ether ester) polyols: synthesis, characterization and properties. *European Polymer Journal* 48, 2097–2106.
- Lai, H.M., Padua, G.W., Wei, L.S., 1997. Properties and microstructure of zein sheets plasticized with palmitic and stearic acids. *Cereal Chemistry* 74, 83–90.
- Lodha, P., Netravali, A.N., 2005a. Characterization of stearic acid modified soy protein isolate resin and ramie fiber reinforced 'green' composites. *Composites Science and Technology* 65, 1211–1225.
- Lodha, P., Netravali, A.N., 2005b. Thermal and mechanical properties of environment-friendly 'green' plastics from stearic acid modified-soy protein isolate. *Industrial Crops and Products* 21, 49–64.
- Manamperi, W.A.R., Fuqua, M.A., Ulven, C.A., Wiesenborn, D.P., Pryor, S.W., 2011a. Effect of canola protein extraction parameters on protein-based plastic properties. *Journal of Biobased Materials and Bioenergy* 5, 500–506.
- Manamperi, W.A.R., Pryor, S.W., 2012. Properties of canola protein-based plastics and protein isolates modified using SDS and SDBS. *Journal of the American Oil Chemists Society* 89, 541–549.
- Manamperi, W.A.R., Wiesenborn, D.P., Chang, S.K.C., Pryor, S.W., 2011b. Effects of protein separation conditions on the functional and thermal properties of canola protein isolates. *Journal of Food Science* 76, E266–E273.
- McCurdy, S.M., 1990. Effects of processing on the functional-properties of canola rapeseed protein. *Journal of the American Oil Chemists Society* 67, 281–284.
- McGregor, D.L., Kimber, D., 1996. *Brassica Oilseeds - Production and Utilization*. European Journal of Lipid Science and Technology 98, 321.
- Mo, X., Sun, X., 2000. Thermal and mechanical properties of plastics molded from sodium dodecyl sulfate-modified soy protein isolates. *Journal of Polymers and the Environment* 8, 161–166.
- Moosavi-Movahedi, A.A., 2005. Thermodynamics of protein denaturation by sodium dodecyl sulfate. *Journal of the Iranian Chemical Society* 2, 189–196.
- Newkirk, R., Canada, C.C.o., 2009. *Canola Meal: Feed Industry Guide*, 4th ed. Canola Council of Canada.
- Rhim, J.W., Gennadios, A., Weller, C.L., Hanna, M.A., 2002. Sodium dodecyl sulfate treatment improves properties of cast films from soy protein isolate. *Industrial Crops and Products* 15, 199–205.
- Schmidt, V., Giacomelli, C., Soldi, M.S., Soldi, V., 2005a. Soy protein isolate based films: influence of sodium dodecyl sulfate and polycaprolactone-triol on their properties. *Macromolecular Symposia* 229, 127–137.
- Schmidt, V., Giacomelli, C., Soldi, V., 2005b. Thermal stability of films formed by soy protein isolate-sodium dodecyl sulfate. *Polymer Degradation and Stability* 87, 25–31.
- Sessa, D.J., Selling, G.W., Willett, J.L., Palmquist, D.E., 2006. Viscosity control of zein processing with sodium dodecyl sulfate. *Industrial Crops and Products* 23, 15–22.
- Shahidi, F., 1990. *Canola and Rapeseed: Production, Chemistry, Nutrition, and Processing Technology*. Van Nostrand Reinhold, New York.
- Solomons, T.W., Fryhle, Craig, B., 2011. *Organic chemistry*. Wiley, Hoboken, NJ.
- Tan, S.H., Mailer, R.J., Blanchard, C.L., Agboola, S.O., 2011. Canola proteins for human consumption: extraction, profile, and functional properties. *Journal of Food Science* 76, R16–R28.
- Zhong, Z.K., Sun, X.S., 2001. Thermal and mechanical properties and water absorption of sodium dodecyl sulfate-modified soy protein (11S). *Journal of Applied Polymer Science* 81, 166–175.

# Review: bio-based films from zein, keratin, pea, and rapeseed protein feedstocks

Weida Shi · Marie-Josée Dumont

Received: 23 July 2013 / Accepted: 28 November 2013 / Published online: 14 December 2013  
© Springer Science+Business Media New York 2013

**Abstract** The development of bio-based films from proteins for their utilization as packaging materials and composites has gained popularity due to the availability, renewable nature, and biodegradability of proteins. The most studied of these polymers is sourced from soy. However, other sources of proteins show great commercial potential when transformed into films. As such, zein from corn and keratin from low-value chicken feather can be plasticized into films which exhibit good mechanical properties. Films derived from pea proteins can resist UV light transmission due to the presence of disulfide bonds within the network and the presence of amino acids such as tyrosine, tryptophan, and phenylalanine. Proteins from rapeseed have not been extensively studied as feedstocks for bio-based films but show great potential since their amino acid composition is similar to soy proteins. This review aims at combining and condensing the body of research performed on the synthesis of zein-, keratin-, pea-, canola-, and rapeseed-based films. The structural diversity of these proteins highly influences their processing methods and the physical properties of the resulting films. The potential applications of these bio-based materials are also highlighted.

## Introduction

The processing of proteins into polymeric matrices is a challenging task due to the complex nature of proteins.

Proteins consist of  $\alpha$ -amino acids joined by peptide bonds. Peptide bonds are amide linkages formed by the condensation reaction of amino acids [1]. These bonds form heteropolymers, which can contain up to 20 different  $\alpha$ -amino acids. The diversity of protein macromolecules depends on the amino acids constituents and the sequence of these amino acids along the chain. Amino acids are characterized by their side chains, which can be electrically charged or uncharged, hydrophobic or polar (Table 1) [2, 3]. Since each sequence may contain any of the 20 amino acids, the number of possible structures may be very large [4]. Different combinations of amino acids result in the formation of various three-dimensional structures.

Protein polymer can be naturally found in different configurations, combinations, or structures [8, 9]. The primary structure is a chain sequence of amino acids [10]. However, some segments within the chain have the tendency to fold into loops, helices ( $\alpha$ -helices), and other simple configurations, which constitute the secondary structure. The secondary structure contains regular substructures which are mainly  $\alpha$ -helix and  $\beta$ -strands [4]. These structures are stabilized by hydrogen bonding. The tertiary structure contains zones of organized and unorganized secondary structure. The biological activities of proteins depend on the tertiary structure which is referred to as the native structure of proteins. The quaternary structure consists of more than one amino acid chain.

Depending on their bioactivities, seed proteins can be divided into two categories, namely enzymes and storage proteins. Enzymes participate in the metabolism of the life cycle of the cell such as in the synthesis of lipids, carbohydrates, and other compounds. Storage proteins are characterized by their ability to store metal ions and amino acids and to provide carbon, nitrogen, and sulfur to support seed germination and seedling growth. They exist in seeds,

---

W. Shi · M.-J. Dumont (✉)  
Department of Bioresource Engineering, McGill University,  
2111 Lakeshore Rd., Ste-Anne de Bellevue, QC H9X 3V9,  
Canada  
e-mail: marie-josee.dumont@mcgill.ca

**Table 1** Characteristics of amino acids [4–7]

Group	Full name	Acronym	Side chain group	Additional side chain characteristics
Nonpolar	Glycine	Gly	–	Nonpolar, carry single hydrogen atom
	Alanine	Ala	Methane	The least hydrophobic
	Valine	Val	Propane	Intermediary hydrophobicity
	Leucine	Leu	2-Methylpropane	Highly hydrophobic
	Isoleucine	Ile	Butane	
	Proline	Pro	Pyrrolidine (Tetrahydropyrrole)	Side chain attach to the $\alpha$ -amino group, forming a pyrrolidine ring
	Methionine	Met	(Methylsulfanyl)ethane	Contain divalent sulfur atom, nonpolar, but capable to weakly interact with polar molecules
Polar neutral	Serine	Ser	Methanol	Contain hydroxyl group
	Threonine	Thr	Ethanol	
	Cysteine	Cys	Methanethiol	Contain thiol group (sulfhydryl)
	Asparagine	Asn	Acetamide	Contain carboxamide, which can serve as hydrogen donor and acceptor
	Glutamine	Gln	Propanamide	
Electrically charged	Glutamate	Glu	Propanoate	Acidic, contain carboxylic group.
	Aspartate	Asp	Acetate	pKa = 4.3 (glutamate), pKa = 3.9 (aspartate)
	Lysine	Lys	Butan-1-amine	Basic, side chain includes nitrogen containing group
	Arginine	Arg	1-Propyl-guanidine	
	Histidine	His	4-Methyl-1 <i>H</i> -imidazole	
	Aromatic	Phenylalanine	Phe	Methylbenzene
	Tyrosine	Tyr	4-Methylphenol	Include phenol ring with hydroxyl group
	Tryptophan	Trp	3-Methyl-1 <i>H</i> -indole	Include indole

roots, shoot tubers, and stem parenchyma of trees and other vegetative organs [11]. Based on their solubility, storage proteins can be divided into four groups; namely the albumins, globulins, prolamins, and glutelins.

Generally, the major storage proteins of cereals in the monocotyledon family are the prolamins and glutelins (also referred as glutenin or glutenine) [11]. The globulins predominate in legumes and oilseed crops, which are in the dicotyledon family, and are accompanied by the presence of a significant amount of albumins in some species [11]. Albumins and globulins are further differentiated according to their sedimentation coefficients into 2S (albumin), 7S (globulin or conglycinin), and 11S (globulin or glycinin) classes [12]. The storage protein contents of seeds, and especially those of cereal, legume, and oilseed crop species, are of great interest in the area of food and biopolymer processing.

This review focuses on the wet and dry processing of proteins sourced from zein, chicken feather, wool, peas, and rapeseed, the physical properties of the resulting polymeric matrices, and their industrial/commercial applications. However, this review does not cover protein-based matrices which are co-polymerized with petroleum-based or biological-based polymers (e.g., polyhydroxyl alkanate and polyhydroxyl butyrate) as excellent reviews on the

topic can be found in the literature [13, 14]. Relationships between the proteins' structures to the processing method and the challenges related to the plasticization of these proteins are discussed. As such, Table 2 provides the molecular weight, glass transition temperatures ( $T_g$ ), and solubility profiles of the proteins discussed in this review [2]. A brief overview about protein-based polymer processing is provided.

## Processing

Usually, low molecular weight plasticizers are added to proteins to facilitate processing as they help in decreasing attractive intermolecular forces, modifying the three-dimensional conformation, and increasing the chain mobility and the free volume [34]. A good plasticizer should demonstrate specific characteristics such as a low melting point and moderate hydrophobicity.

The functional effects of plasticizers on the properties of protein-based polymers include a decrease in rigidity and barrier properties and an increase in flexibility, mechanical resistance, and elongation properties [35–41]. Plasticizers can either be internal or external. The former forms chemical bonds and becomes a part of the final product

**Table 2** Physicochemical properties of proteins [12, 15–33]

Proteins	Main protein fraction	Weight fraction	Molecular weight (kDa)	Solubility	pI	$T_g$ (°C)	Td (°C)	Surface hydrophobicity (S-0)
Zein	$\alpha$ -Zein	80	21–25	80 % Ethanol solution	6.8	164–168	270	$500 \times 10^4$
Keratin feather	$\beta$ -Keratin	91	10	Diluted acidic and basic solution	4.5–5.5	N.A.	201	N.A.
Keratin wool	$\alpha$ -Keratin	95	40–65	Diluted acidic and basic solution	3.22–5.55	160	213	N.A.
Pea protein	Legumin, vicilin	65–80	160–400	Diluted acidic and basic solution	4.3	N.A.	82.1	2734
Canola	Napin and cruciferin	80	14–59	Diluted acidic and basic solution	3.5–5.5	50–65	84 and 102	80

while the latter does not become chemically attached and is removed through evaporation, migration, or extraction [42]. Water is the most used plasticizer; however, it has a low boiling point and may not be efficient at lowering the viscosity of materials when high-temperature processing is required. Therefore, organic plasticizers may be required for high-temperature processing such as extrusion. Note that organic plasticizers should have a low molecular weight and small size in order to penetrate the macromolecular network [43]. Examples of compounds which meet these requirements are polyalcohols (polyols) and urea.

Disulfide bonds present in native proteins are likely to inhibit their flexibility [44]. Therefore, the cleavage of disulfide bonds may be necessary to enable the proteins to unfold to a further degree. Furthermore, it increases the proteins' surface activity, surface hydrophobicity, foaming capacity, and stability [45–47]. Some common reducing agents used for protein processing include dithiothreitol,  $\beta$ -mercaptoethanol, and tri(carboxyethyl)phosphine [48].

Crosslinks, which are bonds that link one polymer strand to another, can create and extend a polymer network. An increase in the degree of crosslinking can enhance the strength, cohesion, rigidity, and water resistance properties of the matrix. Physical crosslinks can be formed with heat treatment or irradiation. Chemical crosslinks are generated by adding agents such as glutaraldehyde, glycerinaldehyde, formaldehyde, tannic acid, lactic acid, and gossypol [49]. The suitability level of crosslinking agents to be used may differ from one type of protein to another.

#### Wet processing

Wet processing consists of solubilizing proteins in large amounts of solvent, casting, and drying the material [50]. The solvent and the pH of the solution are important parameters to consider for a successful processing [51, 52]. Solution casting is generally performed under alkaline

conditions to unfold proteins [53]. As shown in Table 2, it is essential to know the properties of the selected protein source since the solubility of the proteins is a function of their amino acid composition and sequence. The solution casting process is generally slow, requires large volumes of solvents, and thus may be considered commercially less interesting than solid-state or dry processing.

#### Dry processing

Dry processing is a more conventional approach where the proteins are mixed with plasticizers and additives. The mixture is then thermomechanically shaped by compression molding, extrusion, or injection molding. For example, the heat and shear force produced by extrusion can restructure the polymeric matrices in an oriented manner and define the molecular network through dissociation and unfolding of the proteins, which recombine and crosslink afterward [54]. The parameters to be controlled are the mechanical energy input, the shear stress level, and the pressure. Since proteins are heat sensitive, plasticizers must be added to the protein bulk. Otherwise, proteins can undergo degradation as a result of the deformation of the network under heat treatment, which decreases the chain's mobility and thus increases the viscosity of the system. The melting temperature of proteins is above their thermal decomposition temperature [55]. Therefore, the purpose of the plasticizer is to decrease the  $T_g$  of proteins [8, 9]. By decreasing the  $T_g$ , the probability of the proteins to degrade is reduced as the viscosity will decrease and the chains' mobility will increase. Therefore, the  $T_g$  should be adjusted below the decomposition temperature. It is worth mentioning that  $T_g$  is not a native property, it is affected by several factors, including molecular weight, polarity of the lateral groups and chain rigidity, the presence of crystalline zones or intermolecular bonds, and the type of plasticizer being used and its concentration [56, 57].

## Corn zein-based films

Recently, the FAO reported that the annual global production of corn was around 840 million metric tons [58]. Corn contains a prolamine protein, zein, which is exclusively found in corn [59]. The corn kernel contains more than 60 wt% of starch and has a protein content ranging from 7 to 12 wt% [60] of which zein accounts for 44–79 wt% overall [61, 62]. Zein is mainly isolated from corn gluten meal, the by-product of corn wet milling.

As shown in Table 2, zein is insoluble in water, unless alcohol or alkali (no less than pH 11) or high concentration of urea or anionic detergents or acid (less than pH 2) is present in the solution [60]. The insolubility of zein partially results from its amino acids composition (Table 1), which mainly consists of glutamic acid (21–26 % wt%), leucine (20 wt%), proline (10 wt%), and alanine (10 wt%) [60]. The high fraction of nonpolar amino acid residues affects the solubility behavior of zein to some extent. Average molecular weight of zein ranges from 9000 to 26000 Da [59]. Four distinct types of zein ( $\alpha$ -,  $\beta$ -,  $\gamma$ -, and  $\delta$ -zein) have been classified based on their molecular weights and solubility in aqueous alcohol and reducing agents.  $\alpha$ -Zein is the most abundant type and it is the predominant fraction present in commercial zein [63, 64].

### Wet processing

Zein has been recognized for its film-forming properties for a long time. A plethora of plasticizers with a polarity ranging from polar to nonpolar have been tested. The use of hydrophilic plasticizer may increase the water content of the matrix as it can facilitate water absorption from air, which confers flexibility to the matrices at low relative humidity (RH) as compared to hydrophobic plasticizers. This characteristic may become a severe inconvenience when matrices are exposed to high RH, as the increase in water content renders the matrices very weak [65]. On the other hand, hydrophobic plasticizers decrease the water content of zein matrices and increase their rigidity [65, 66]. The physical properties of films with different common plasticizers were measured. The amount of water absorbed by the films was found to follow, to a certain extent, a logical order based on the hygroscopy of the plasticizer: glycerol > triethylene glycol > levulinic acid > PEG 300 > none > dibutyl tartrate > oleic acid [65]. Great caution must be taken while measuring water absorbance and water vapor permeability (WVP) when unsaturated fatty acids are used as plasticizer since the measurements are dependent on the temperature. For example, samples formulated with oleic acid showed improved water barrier property at low temperature as compared to room temperature due to the crystallization of the fatty acid [67].

The literature reports that combining plasticizers within the same polymeric system can greatly improve the properties of the resulting matrices. Xu et al. [68] investigated the plasticizing effects of oleic acid and glycerol. Oleic acid is amphiphilic and can embed in the protein matrix and increase the internal chain movement. This is referred to as internal plasticization or molecular plasticization. But hydrophilic glycerol molecules can only place themselves within the hydrophilic region of the proteins, resulting in an increased number of side chains resulting in structural plasticization [69, 70]. The synergetic interactions between oleic acid and glycerol were reported to improve the mechanical properties of films as compared to glycerol alone [68]. The elongation at break of films plasticized by the combination of glycerol and polypropylene glycol (PPG) reached 117.8 %. Such improvement was attributed to the synergy between plasticizers resulting especially from the secondary forces exerted by the hydrophobic methyl side chain on PPG. This changed the protein structure and exposed hydrophilic portions to the surrounding glycerol to further reduce the intermolecular forces between the protein molecules [71]. The WVPs of 1.01, 1.06, and 0.62 (g mm kPa<sup>-1</sup> hr m<sup>2</sup>) were reported for zein-based films formulated with 15 wt% of glycerol, 30 wt% of glycerol/PPG, and no plasticizer, respectively. Films without plasticizer had nearly doubled the WVP according to the result.

Preparation of films from zein resin rather than from aqueous solution was reported as an efficient method to increase the mechanical properties of the matrices. Lai et al. [72] prepared resin films by forming a zein–oleic acid emulsion. The soft moldable resin was obtained by precipitation of the emulsion using cold water. Thereafter, films were formed by stretching the resin over a rigid frame followed by a drying step [72]. The films were reported to exhibit better mechanical properties than zein cast films also plasticized with oleic acid [72]. As such, the elongation at break, modulus, and toughness of resin films were 12 %, 267, and 0.8 MPa, respectively, as compared to 3 %, 317, and 0.08 MPa, respectively, for cast films [72]. Analysis by differential scanning calorimetry showed that oleic acid separated from zein cast films during heating probably accounted for their brittleness [72]. On the contrary, no phase separation occurred in resin-based films which resulted in an increased toughness [72]. The same group of authors also studied the moisture absorption of resin films. It was shown that applying a drying oil coating (flax and tung) followed by curing conditions under UV light or  $\gamma$ -radiation reduce moisture absorption [73]. These films were insoluble in water and exhibited superior mechanical performance than zein cast films. UV light and  $\gamma$ -radiation promoted oxidation and speeded up the polymerization process [73].



Another efficient method to prevent phase separation is to chemically modify zein with a crosslinker. Zein has been crosslinked with glutaraldehyde, formaldehyde, citric acid, epochlorohydrin, 1,2,3,4-butanetetracarboxylic acid, 1,2-epoxy-3-chloropropane, dialcohols, and polymeric dialdehyde starch among others to improve the mechanical properties of the resulting matrices [71, 74–76]. Two mild crosslinking agents (1-[3-dimethylaminopropyl]-3-ethylcarbodiimide hydrochloride and *N*-hydroxysuccinimide) were reported to suppress aggregate formation in the films and to improve their mechanical properties due to the formation of intermolecular amide bonds [77]. Acylation of zein with lauryl chloride was shown to be an efficient way to improve the overall properties of zein films [78]. Since the plasticizer is chemically attached to zein, no micro-phase separation was detected in the matrices. Furthermore, lauryl chloride was proven to efficiently plasticize zein since the  $T_g$  decreased by almost 26 °C. Additionally, this modification increased the elongation of the films by sevenfold. Derivatization of zein with octenyl succinic anhydride, alkyl ketene dimer, and alkenyl ketene dimer has also proven to be an efficient method of changing the surface morphology of the films and of decreasing their water absorbance capacity [79].

Recently, the effects of incorporating plasticizers with phenolic acids or flavonoids, such as gallic acid, *p*-hydroxybenzoic acid, ferulic acids, catechin, flavone, and quercetin were investigated [80]. Compared to plasticizing with glycerol, incorporation of phenolic acid could not only reduce the brittleness and increase the flexibility, but also provide zein films with antioxidant and antimicrobial characteristics [80]. Incorporation of other agents such as thymol [81], an essential oil found in thyme, or nisin [82] also conferred antibacterial properties to zein films.

#### Dry processing

There are several factors which influence the physical properties of batch mixed or extruded zein. The processing temperature has clearly a direct impact on the properties of the materials since proteins are sensitive to heat [83, 84]. The choice of a single-screw extruder versus a twin-screw extruder also has an impact on the resulting properties. Twin-screw extruders exert a more thorough mixing action, which results in superior moisture loss from the protein as compared to single-screw extrusion. The lower content of moisture caused a decrease of the films' elongation at break. However, the microstructure of the extrudates is more compact since there are less void spaces. This has a positive effect on the tensile strength of the materials [85]. Attention must be given to the orientation of the sample when testing for mechanical properties since discrepancies were found when samples were tested in the machine

direction versus the transversal direction. This is due to the materials' anisotropic behavior. For example, Selling and Sessa [86] found that zein ribbons tested for elongation in the machine direction were 33–291 % higher than when tested in the transverse direction. This tendency was also true for tensile strength where values were 18–64 % superior in the machine direction.

Interestingly, the molecular structure of zein was demonstrated to have an impact on the processing method. The molecular structure varies as a function of the corn source from which it is isolated from, the extraction techniques used, and other factors [87]. The content in  $\alpha$ -helix and  $\beta$ -sheets can thus vary greatly. As the ratio of  $\alpha$ -helix to  $\beta$ -sheets increases, the diffusion of plasticizer inside the protein improves, since less aggregated secondary structures are present. Oliviero et al. [87] found that zein with high  $\alpha$ -helix content could be processed by film-blowing technology, and that this was not possible for zein containing a high content of  $\beta$ -sheets. The ratio of  $\alpha$ -helix to  $\beta$ -sheets was also discovered to change during the mixing process. The amount of  $\beta$ -sheets tends to decrease, favoring a less ordered structure and, therefore, improving mechanical properties [87].

Similar to wet processing, various plasticizers, solvents, additives, and methods were utilized for the preparation of zein films through dry processing. Several studies were done on the properties of moldable resins plasticized with various fatty acids and processed by rolling, extrusion, or hot pressing [83, 85, 88, 89]. Santosa and Padua [88] used oleic acid and linoleic acid as plasticizers and compared their effects on the mechanical properties of the resulting zein sheets. Both fatty acids could increase the elongation at break and the toughness of the sheets, but the tensile strength values were improved when oleic acid was utilized. It was hypothesized that oleic acid facilitated the structural mobility of the matrix. On the other hand, the addition of linoleic acid was found to increase the water barrier properties of the sheets compared to the addition of oleic acid. The authors explained the difference by the polymerization of linoleic acid during sheet preparation, which may have sealed off pores and gaps at the surface of the sheets [88].

Glyoxal was also used as crosslinking agent to improve the physical properties of thermoplastic zein, as it modified the structure by reacting with the functional groups of lysine, arginine, and the N-terminal amine in solution or in the melted state [79, 84, 90]. Similar behavior was observed for citric acid and 1,2,3,4-butanetetracarboxylic acid [84, 90]. Therefore, the addition of glyoxal in concentrations varying from 1 to 4 wt% increased the tensile strength and elongation at break from 24 to 35.2 MPa and from 4 to 6.4 %, respectively, in injection-molded matrices [79]. However, when tested on samples that have been

**Table 3** Physical properties and conditioning of thermoplastic zein material

Preparation method	Form	Plasticizer	Solvent	Additives	Tensile strength (MPa)	Elongation at break (%)	Modulus (MPa)	References
Hot pressing	sheets	50–100 % Oleic acid/ linoleic acid	80 % Isopropyl alcohol	None	1.6–9.4	5.9–165	49.9–557.3	[88]
Twin-screw extrusion and molding	Bar	5–10 % TEG	None	Glyoxal and NaOH	15.4–35.2	4–30.6	305–1126	[79]
Pressing	Bar	None	None	Glyoxal, NaOH KOH, and Ca(OH) <sub>2</sub>	28.9–40.6	8.7–14.1	379–451	[84]
Twin-screw extrusion	Sheets	Palmitic, stearic, oleic, linoleic and linolenic acid	70 % Ethanol	None	6.6–11.6	0.8–25.5	311–1967	[83]
Hot pressing	Slabs	25 % TEG 400	None	None	0.04–3.57	42–270	4.1–383	[87]

compressed, no increase was obtained for tensile strength and elongation at break values [79]. No explanations were provided.

As outlined at the beginning of “[Corn zein-based films](#)” section, zein is insoluble in water but can be solubilized by addition of an alkaline agent. Interestingly, the choice of the alkali was shown to have an impact on the properties of the resulting matrices. Woods et al. studied the effects of the addition of NaOH, KOH, and Ca(OH)<sub>2</sub> on the physical properties of zein films (Table 3). The addition of KOH and Ca(OH)<sub>2</sub> decreased the tensile strength and elongation at break (29 MPa and 10 %, respectively) as compared to the addition of NaOH (40.6 MPa and 14.1 %, respectively) when all bases were added at the 0.0186 mmol of OH<sup>-</sup> per gram of zein. The authors did not provide any explanation [84].

To produce the edible grade film, it is prohibited to use any toxic chemicals. Sugars are widely available and can act as natural plasticizers. For example, the mechanical properties of films plasticized with galactose, fructose, and glucose were compared. The tensile strength, elongation at break, and modulus of galactose-plasticized film (0.5–1.0 g plasticizer/g zein) ranged from 18 to 20 MPa, 3 to 3.5 %, and 1050 to 1350 MPa, respectively [91]. Hot-pressed films plasticized with galactose had superior tensile strength and elongation at break than films plasticized with other sugars [92]. This was explained by the fact that fructose and glucose are more hydrophilic than galactose. Therefore, films plasticized with fructose and glucose were more prone to absorb water, which resulted in their lower mechanical properties [92]. But the tensile properties of zein films plasticized with sugars are generally not as good as using conventional plasticizers, especially in terms of elongation at break.

## Applications

Zein has great potential for use in food packaging due to its antioxidative nature [93, 94]. This character allows zein-based coatings to have antioxidative function without adding any antioxidant during processing. Some studies have report benefits of using zein films as packaging material for cooked turkey and fresh broccoli [95, 96]. In addition to food packaging, the application of zein in the tissue engineering area has also been investigated. In this context, zein films were found to exhibit good biocompatibility and suitability to be used as scaffolding for tissue engineering, i.e., as adhesive substrates for implanted cells and as physical supports and guide for the formation of new organs [67, 97]. Earlier studies showed that zein films could also be used for controlled drugs delivery [98, 99]. Edible film made from the blending of zein and other additives is reported to be used as an oral flavor-masking agent [100].

## Keratin

Keratin is a family of fibrous protein, which is mainly found in hair, nail, epidermis, hoof, horn, and feather. Keratin can be found in two conformations:  $\alpha$ -keratin or  $\beta$ -keratin [18]. The  $\alpha$ -keratin is mainly found in mammalian epithelia, such as hairs, horns, nails, hooves, and claws [4] while  $\beta$ -keratin, also called hard protein, is mainly found in feathers, claws, beaks of birds, nails, scales of reptiles, and quills of porcupines [101, 102]. There are 14 types of amino acids found in the composition of keratin [18]. Keratin is characterized by its content in the amino acid cysteine, which varies from 2 to 18 wt% [103]. The

presence of cysteine results in high strength and resistance against the action of proteolytic enzymes due to the disulfide bonds between cysteine molecules [103]. Keratin is also characterized by its high content of the hydroxyl amino acid, serine, which accounts for 15 wt% [103].

Chicken feathers are low-value agricultural products where the content in keratin varies from 85 to 99 wt% [104, 105]. A small proportion is used for the production of low-value animal feed, but it is mostly discarded. There are 1.13 million tonnes of dry chicken feathers produced in the US and 5 million tonnes of feathers produced globally each year [106, 107]. A few researchers have been investigating the potential of using keratin found in chicken feathers and wool as a substitute for petroleum in the synthesis of polymers. The following sections detail their findings.

### Wet processing

Keratin must be extracted from its source (e.g., chicken feather or wool) prior to its utilization. Compared with other fibrous proteins, such as elastin, collagen, and myofibrillar proteins, the distinctive character of keratin is due to its high content of cysteine [108]. This amino acid is characterized by its extensive crosslinking and highly hydrophobic properties, resulting in its insolubility in either polar or apolar solvents [108]. To dissolve keratin, the disulfide and hydrogen bonds must be broken. This can be achieved through alkaline and acid hydrolysis, by a combination of enzymatic and chemical treatments, or by the use of ammonium copper hydroxide [109–111]. The literature reported the preparation of stable keratin dispersions through the use of 2-mercaptoethanol (2-Me) in combination with urea and sodium dodecyl sulfate (SDS) in alkaline solution [108, 112, 113]. This extraction method was characterized by a high keratin yield (75 %) and undamaged keratin extracts [105, 108, 112–114]. However, this method was criticized due to its industrial infeasibility due to the high cost of 2-Me [115]. The use of sodium sulfide (Na<sub>2</sub>S) as the sole digestion agent was proposed by Poole et al. [115]. A yield of 62 % was reported through this method. Interestingly, it was shown that the extraction method has an influence on the properties of the resulting films (Table 4). From this table, the films synthesized from keratin digested by Na<sub>2</sub>S showed superior tensile properties than by other methods.

Knowledge of the strain at break values would have been interesting if the authors had plasticized the matrices. Stable keratin dispersions were also prepared from wool using the 2-Me and SDS digestion method [113]. The superior ductility was attributed to the high glycerol content.

The choice of plasticizer also greatly influences the properties of the films. The use of sorbitol, glycerol, and PEG had different effects on the microstructure, sorption isotherm, and WVP properties of the films [114]. Glycerol and sorbitol, due to their low molecular weight, can easily diffuse into the protein dispersion as compared to PEG, which led to matrices with more uniform structures. The low molecular weight of glycerol also facilitated its integration within the keratin chains and, therefore, its association with hydrophilic amino acids such as tyrosine and serine [114]. Consequently, films processed with glycerol had the highest monolayer moisture content (measure of water sorption ability) followed by sorbitol and PEG [114]. The moisture content of monolayers increased with an increase in plasticizer concentration. This resulted in a decrease in the film’s barrier properties [114]. For example, the WVP of the films increased from  $3.46 \times 10^{-4}$  to  $3.84 \times 10^{-3}$  g mm m<sup>-2</sup> h<sup>-1</sup> kPa<sup>-1</sup> as the concentration of glycerol increased from 0 to 0.09 g/g keratin. An explanation was provided based on the mobility of the polymeric chains. The addition of a plasticizer resulted in an increase in the mobility of the chains and a decrease in density. This allowed for the absorption of water molecules. It was demonstrated that the brittleness of the films could be reduced by addition of PEG without affecting the WVP due to its enhanced hydrophobic character [105]. Another efficient method to increase waterproofness is to induce chemical crosslinking by the addition of ethylene glycol diglycidyl ether and glycerol diglycidyl ether [117].

### Dry processing

Keratin-based polymers are considered as thermosetting polymers due to the high content of cysteine bonds. Generally, proteins containing high amounts of mobile hydroxyl groups are easier to process by dry processing than those containing high amounts of immobile cysteine bonds [106]. Keratin sourced from chicken feather is composed of 22 % of hydroxyl amino acids (serine, threonine, and tyrosine) and 7 % cysteine [106], while in wool it is 20 and

**Table 4** Mechanical properties of keratin-based films processed with different extraction and digestion methods

Digestion method	Keratin source	Plasticizer	Tensile strength (Mpa)	Strain at break (%)	References
Na <sub>2</sub> S 1 h	Chicken feather	None	61	7	[115]
2-Me		13 % Glycerol	30	3	[112]
2-Me and SDS		None	17	2	[116]
2-Me and SDS	Wool	50 % Glycerol	0.25	42	[113]

11 %, respectively. This suggests that feather keratin is more processable than wool keratin. Because of the high content of cysteine, redox agents (e.g., sodium sulfite) may be necessary to disrupt the disulfide bonds between the cysteine residues in order to reduce the blend viscosity and increase the processability. Barone et al. [106] studied the processability of keratin as a function of the concentration of sodium sulfite. Interestingly, they found that the lowest viscosity occurred when the concentration of sodium sulfite was at 3 wt%, above which the viscosity increased. This was explained by the greater accessibility of intermolecular disulfide bonds than intramolecular disulfide bonds. At a concentration of 3 wt%, all the intermolecular disulfide bonds were reduced. Above that concentration, the extra sodium sulfite started to reduce the intramolecular disulfide bonds, resulting in a rearrangement of keratin and causing its crystallization upon oxidation. Consequently, an increase in crystallinity increased the viscosity [106]. Furthermore, Barone et al. also found a correlation between the granular size of the keratin and the diffusion of sodium sulfite within the keratin. Smaller sized particles have larger surface areas, which creates more reaction sites. This allowed for an improved diffusion of sodium sulfite into the solid. As a result, as the initial size of keratin increased, both stress and strain at break points of the resulting matrices decreased [106].

In addition, the cysteine content of keratin-based polymers greatly influences their ability to biodegrade. Barone and Arikian studied the biodegradation of extruded keratin plasticized with 37.5 wt% glycerol and containing 3 wt% sodium sulfite. Surprisingly, only 24 % of the matrices degraded in 30 days while soy, wheat, whey, and collagen-based matrices completely degraded [118]. The limited biodegradation was attributed to the lack of microbes producing keratinase to degrade the disulfide bonds [118]. Even though sodium sulfite was added to cleave the disulfide bonds, the cysteine residue could still stabilize the keratin structure which resulted in a partial degradation.

The quality of keratin films also depends on the processing parameters and the conditioning environment. Interestingly, it was found that the moduli of thermoprocessed keratin films decreased as the samples were pressed for longer periods of time [106]. Longer processing times also led to a slight decrease in tensile strength and an increase in elongation at break. This was partially explained by the nonlinear viscoelastic nature of the films. Furthermore, the melting point of the films was 241 °C after pressing. After transferring the films into an oven at 60 °C for 21 h, the melting temperature increased to 248 °C. This was explained by the annealing process which started at 60 °C. This modified the crystal structure of the films which translated into an increase in tensile strength and moduli.

Structural transformation from  $\alpha$ -helix to  $\beta$ -sheet occurred as a result of keratin hot pressing in the presence of plasticizer under high temperature and pressure [106]. These structural changes were not observed for extruded keratin.

#### Potential applications

Keratin from feather meal can be obtained at low cost and can be used to make films [119]. The main applications for keratin-based films are in the manufacture of agricultural films, compostable packaging, and edible films and coatings [120, 121]. It was also found that keratin cast films are a suitable medium for the attachment and proliferation of mouse fibroblast cells, which suggested their use in biomedical applications [122, 123]. Good mammalian cell adhesion and proliferation on wool keratin films were also reported [124]. This finding suggests the promising use of keratin film in the tissue engineering and medical fields [64, 107].

#### Pea protein

Pea (*Pisum sativum*) is a cool-season pulse crop. Its annual worldwide production is 10–12 million tonnes [125]. Peas were traditionally used for animal feed, especially for swine, due to their high nutritive value [126]. But as the price of peas increased in recent years, they are now mainly purchased for human consumption or high-value feeds (pet food or aquaculture). Peas can be divided into two types: (1) green “garden” peas (*P. sativum hortense*) and (2) field peas or dry peas (*P. sativum* ssp. *arvense*) [127]. Fresh or frozen garden peas are usually consumed as vegetables, and dry peas are also used for the production of protein and starch.

The protein content of dry peas ranges from 14 to 39 wt% [25] and contains mainly globulin, which account for 65–80 % of the protein [128]. The globulin fraction contains legumin (MW 360–400 kDa), vicilin (MW 160–200 kDa), and small quantities of convicilin (MW 280 kDa) [25, 129, 130]. As for the amino acid profile, pea protein mainly contains lysine (8 %), arginine (8 %), aspartic acid (12 %), and glutamic acid (19 %) [131]. It is also characterized by a low content of methionine, cysteine, and tryptophan [131].

Pea proteins can be processed into pea flour, pea protein concentrate, and pea protein isolates (PPI) [25, 132, 133]. Because of its high nutritive value and non-allergic character, PPI is mostly used as an additive to enrich the protein content in the food industry. The number of technical applications of pea protein is very limited including surfactants, films, and microspheres in cosmetics [134].



**Table 5** Mechanical properties of plasticized films in function of the plasticizer chain length [137]

Plasticizer	Strain at break (%)	Tensile strength (MPa)
Ethylene glycol	152	2.2
Diethylene glycol	137	2.1
Triethylene glycol	121	1.0
Tetraethylene glycol	100	0.7
Glycerol	75	0.5

### Pea protein-based films

The processing temperature was demonstrated to have an impact on the film-forming mechanisms of pea proteins [135, 136]. As pea proteins are being denatured by heat, the formation of intermolecular  $\beta$ -sheets is being induced, while the amounts of  $\alpha$ -helix and random coils are decreasing [137]. The denaturation also decreases the water binding capacity and the water–protein association [135]. The reduction of water within the films should normally reduce the ductility of the material since water acts as a plasticizer. However, it was shown that the tensile strength and elongation at break values increased with heat denaturation. The improvement in tensile strength was explained by the formation of intermolecular disulfide bonds [135]. Furthermore, as water is eliminated, plasticizers interact through hydrogen bonds with the newly formed  $\beta$ -sheets [137]. These interactions were found to be stronger for plasticizers with short aliphatic chains, and resulted in improved mechanical properties (Table 5) [137, 138]. As shown in Table 5, films plasticized with ethylene glycol exhibited the highest values of elongation at break (152 %) and tensile strength (2.2 MPa) due to the formation of strong hydrogen bonds formed between the hydroxyl group of ethylene glycol and the adjacent polypeptides [137]. It is generally believed that the shorter the aliphatic chain, the stronger the hydrogen bonds. Therefore, as the chain length increased, the interaction becomes weaker [137]. Although, the chemical formula of glycerol is similar to that of ethylene glycol and diethylene glycol, the interactions between glycerol and the polypeptides are more complex due to its three hydroxyl groups, which lead to the formation of a less organized network with poorer physical properties [137]. However, no clear explanation was provided for this.

Since pea protein films are hydrophilic, the addition of monoglycerides, which are lipophilic molecules, should decrease their WVP. This was studied by the addition of monoglycerides of heptanoic, undecanoic, oleic, linoleic, liolenic, and erucic acid to films plasticized with diethylene glycol and they did not decrease the WVP of the resulting films but did allow very soft films to be formed, improving their aging behavior [138]. Further attempts to decrease the

hydrophilicity of the films were performed. In that optic, Kowalczyk and Baraniak [139] studied the properties of pea proteins synthesized under neutral and alkaline conditions (pH 7.0, 9.0, and 11.0). Glycerol and sorbitol were chosen as plasticizers [139]. No difference in WVP was observed at the various pH values. As for the plasticizers, the use of sorbitol decreased the WVP compared to films plasticized with glycerol. This was due to the enhanced hydrophilic nature of glycerol compared to sorbitol. Furthermore, heat treatment had no influence on the WVP of the films but increased their transparency as compared to non-heated films. This was attributed to the formation of a more stable network due to heat denaturation [139]. The lowest WVP of pea protein-based film that can be found is  $19 \text{ g m}^{-1} \text{ s}^{-1} \text{ Pa}^{-1}$ , which was achieved by the pea protein-based film plasticized with 9.2 wt% of ethylene glycol [138]. It was also found that the films resisted UV light transmission, which resulted from the presence of UV-absorbing chromophores induced by disulfide bonds and amino acids such as tyrosine, tryptophan, and phenylalanine. From this particularity, the authors suggested that pea protein films could be used as packaging material to prevent degradation from UV light. This could potentially lead to applications in food packaging [139].

### Rapeseed proteins

Rapeseed is an important player in the vegetable oil industry. It is now the second largest oilseed crop after soybean and it is grown extensively in India, China, Canada, and northern Europe [140]. Four major rapeseed species are grown: *Brassica rapa* (turnip rape), *B. napus* (swede rape), *B. juncea*, and *B. carinata* [140]. The erucic acid content found in rapeseed ranges from 22 to 60 % [141]. Erucic acid is known to cause damage to human hearts and this had compromised the nutritional value of rapeseed oil for years. In 1968, a genetically modified rapeseed variety containing low erucic acid was produced in Canada [141]. Canola, a rapeseed species *B. napus* L, is now the third most widely grown commercial genetically modified crop after soybean and maize with an area equivalent to that of cotton [140]. Canola contains low level of erucic acid (less than 2 %) and glucosinolates (less than  $30 \mu\text{mol g}^{-1}$ ) [141, 142]. The Canadian production of canola reached 14.2 million tonnes in 2011 [143]. The main components of canola are oil (40–43 %) and proteins (17–36 %) [144, 145]. Canola meal, the by-product of oil extraction, is currently used as protein source in low-value animal feed [142]. In recent years, canola protein concentrates and isolates have been considered as an alternative source of plant protein for various applications because of its well-balanced amino acid composition; in terms of



**Table 6** Amino acid profile of canola protein and soy protein (value is expressed in percentage)

Amino acids		Defatted canola meal (%)	Canola protein isolates (CPI) (%)	Soybean (%)	Soy protein isolates (%)	Soy protein concentrates (%)
Essential	Histidine	2.57	2.45	2.77	2.46	2.49
	Isoleucine	2.82	3.03	4.97	4.69	4.49
	Leucine	7.22	7.52	7.81	8.01	7.57
	Lysine	5.62	4.74	6.98	6.23	6.22
	Methionine	1.74	1.91	1.38	1.29	1.3
	Cysteine	0.77	0.63	1.45	1.26	1.49
	Phenylalanine	3.84	4.36	5.41	5.11	4.97
	Tyrosine	2.88	3.02	3.43	3.81	3.66
	Threonine	4.54	4.13	4.4	3.66	3.63
	Valine	3.66	3.77	5.25	5.02	4.715
Non-essential	Alanine	4.71	4.66	4.66	4.22	4
	Arginine	5.38	6.32	7.91	7.9	8.02
	Aspartic acid	7.82	8.32	12.8	11.84	12.35
	Glutamic acid	17.75	17.28	20.47	19.25	20.04
	Glycine	5.25	5.38	4.57	4.09	4.11
	Proline	5.82	5	6.01	5.24	5.28
	Serine	4.81	4.71	5.6	4.98	5.04
Reference		[153]	[153]	[153, 154]	[155]	[155]

which, canola protein is very close to soy protein (Table 6). Both canola and soy proteins contain a high content in glutamic acid, aspartic acid, and leucine and a low content in cysteine, methionine, and histidine. The average molecular weight of canola protein isolate ranges from 14 to 59 kDa, depending on the method of extraction [24]. Canola protein is mainly composed of two major storage proteins: napin (albumin) and cruciferin (globulin), accounting for 20 and 60 % of the total protein content, respectively [146]. Napin is a 2S albumin. It is characterized by its strong alkaline character, which is attributed to a high level of amidation of amino acids. Cruciferin is a neutral 12S globulin and is featured by a high molecular weight (300–310 kDa). It usually acts as a gelling agent in its native form [147]. Even though canola proteins possess a wide range of isoelectric points (pI) from pH 4 to 11, the major pI of canola proteins lies between pH 3.5 and 5.5 [148–151]. The thermal denaturation and enthalpy of denaturation values of canola protein were reported at 88.12 °C and 6.19 J g<sup>-1</sup>, respectively [152].

The commercial uses of canola protein in the food industry have been limited due to the presence of antinutritional compounds such as phenolics, glucosinolates, and phytates [23, 156]. Under alkaline conditions, phenolic compounds, which are highly reactive molecules, undergo both enzymatic and non-enzymatic oxidation to form quinones. Quinones react with proteins to generate a dark green/brown color in the products [147, 157, 158]. In addition, phenolics are characterized by a bitter flavor,

which is an undesirable quality for food products. More importantly, studies performed on animals revealed the antinutritional and the toxic character of glucosinolates [142]. It was shown that these molecules can interfere with the thyroid which results in growth reduction [159]. Finally, phytates were reported to react with minerals and form phytic acid–mineral complexes, resulting in a decrease in mineral metabolism [142, 160, 161]. These antinutritional compounds are found when canola proteins are extracted under alkaline conditions, which is the most common method of extracting plant proteins. The concentration of these compounds can be reduced by using either other extraction methods such as the protein micellar mass method, or through membrane technology [148, 162].

Compared with soy protein, wheat gluten, and corn zein; rapeseed proteins have a lower commercial value due to the presence of antinutritional compounds. Therefore, the use of canola/rapeseed protein-based polymers for non-edible applications should be economically competitive.

#### Rapeseed protein-based films

There are very few studies on the synthesis of rapeseed protein films [163, 164]. In one of these studies, gelatin and *Gelidium corneum*, a type of red algae, were blended with rapeseed protein separately to produce films plasticized with sorbitol or sucrose through solution casting [163]. It was found that stronger intermolecular interactions occurred between the rapeseed proteins and gelatin molecules

**Table 7** Comparison of physical properties of films from standard CPI and refined CPI [151]

	Plastic from standard CPI	Plastic from refined CPI
Tensile strength (MPa)	10.3	11.9
Tensile modulus (Mpa)	192	200
Elongation (%)	17.5	23.5
Water absorption (%) 24 h	6.6	5.9
Glass transition temperature (°C)	−21.9	−22.9

which resulted in improved physical properties [165]. In another study, rapeseed proteins were successfully chemically modified through acetylation, succinylation, and phosphorylation to change their solubility, surface tension, and foaming stability. These chemical modifications could potentially improve the physical properties of the resulting rapeseed protein films [166]. However, based on the literature available, no film was synthesized from the chemically modified rapeseed proteins.

The only reported application regarding rapeseed protein-based films was as packaging films for strawberries. In fact, the paper was more focused on the addition of grape seed extract in the formulation since it was proven to confer an antimicrobial character to the films [167].

The literature reports films made from canola proteins and synthetic polyester by reactive extrusion. The weight ratio of protein to synthetic polyester was 8:7 [151, 168, 169]. The composition of canola protein isolates was reported to be an important factor influencing the properties of the films [151, 168]. As stated above, the pI of canola proteins lies between pH 4 and 11. Even though most of the protein fractions are isolated at or below pH 7, napin precipitates at pH 11 [151, 168]. Napin is characterized by its high hydrophilic character. Films made without the napin fraction (refined isolates) showed improved mechanical properties and water resistance compared to films that included all protein fractions (standard isolates) (Table 7). The refined protein isolates showed improved tensile properties due to the increase in hydrophobicity caused by the elimination of most of the hydrophilic fraction. Consequentially, the compatibility and interaction between the proteins and the synthetic polyester were enhanced, leading to an increased number of interlocks and entanglements between the protein–protein and protein–polymer molecules [168]. The refined protein isolate films were featured by a more compact structure, which is also evidence of a higher degree of interactions between synthetic polyesters and the hydrophobic protein isolates [151]. This resulted in a moderate

reduction of water absorption, decreasing from 6.6 (standard isolates) to 5.9 % (refined isolates) [151].

It was shown that the addition of SDS and sodium dodecyl benzene sulfonate (SDBS) denaturing agents could greatly modify the properties of the resulting films. As compared to unmodified protein-based matrices, the addition of SDBS at a concentration of 5 % increased the water absorption by up to 115 %. On the contrary, the addition of 5 % of SDS decreased the water absorption capacity by 77 % [169]. The results generated by the addition of SDBS were explained by the structure of the denaturing agent. It was hypothesized that  $\pi$ – $\pi$  bonds formed by the interaction of the SDBS' benzene ring with water leads to a negative charge delocalization. These structural changes were not observed for SDS molecules. On the other hand, the addition of both SDS and SDBS increased the fat absorption of canola proteins. The authors state that this could potentially be beneficial for food-based applications [169].

Canola protein is a novel source for manufacturing the polymer film; therefore, no commercial application of the polymer film made from canola protein alone has been recorded yet. However, canola protein and other plant protein blending with starch and algae is shown to be an alternative source for manufacturing the thermoplastic film with strong absorption capacity, which may be used as a personal hygiene product [170].

## Conclusions

This review presented the potential for proteins sourced from agricultural by-products or considered as marginal to be transformed into bio-based polymers. As was reviewed, zein, keratin, pea, rapeseed, and canola proteins differ in their amino acid composition which greatly affects the processability of the proteins and, therefore, the physical properties of the resulting polymeric matrices. Processing zein can be challenging since zein is insoluble in water due to its amino acids composition. Furthermore, it was shown that the content in  $\alpha$ -helix and  $\beta$ -sheets can greatly vary depending on the source of the zein and the extraction techniques which leaves the structure of zein with more or less aggregated secondary structures. It was reported that zein-based films could be used in food packaging due to its antioxidative nature. Processing keratin is complex as it contains a high ratio in cysteine, which is characterized by its extensive crosslinking and highly hydrophobic properties. The content in cysteine greatly limits the biodegradation of keratin-based polymeric matrices. Keratin-based films are suitable in the manufacture of agricultural films, compostable packaging, and edible films and coatings. Proteins from peas are characterized by their low content of

cysteine, rendering these proteins easier to process. Pea protein films are characterized by their resistance to UV light transmission due to the presence of UV-absorbing chromophores induced by disulfide bonds and amino acids such as tyrosine, tryptophan, and phenylalanine. This UV protection could be a serious advantage for applications in food packaging.

**Acknowledgements** The financial support of McGill University through the Start-Up Fund is gratefully acknowledged. We would also like to thank Dr. Darwin Lyew and Ms. Elizabeth Kendrick for editing help.

## References

- Solomons TWG, Fryhle CB (2011) Organic chemistry. Wiley, Hoboken
- Bastoli C, Rapra (2005) Handbook of biodegradable polymers. Rapra Technology, Shawbury
- Nölting B (2006) Protein folding kinetics biophysical methods. Springer, Berlin
- Kessel A, Ben-Tal N (2011) Introduction to proteins: structure, function, and motion. CRC Press, Boca Raton
- Iwaoka M, Takemoto S, Okada M, Tomoda S (2002) Weak nonbonded  $s \cdots x$  ( $x=O, N,$  and  $S$ ) interactions in proteins. Statistical and theoretical studies. Bull Chem Soc Jpn 75:1611–1625
- Jordan IK, Kondrashov FA, Adzhubei IA et al (2005) A universal trend of amino acid gain and loss in protein evolution. Nature 433:633–638. doi:10.1038/nature03306
- Burley SK, Petsko GA (1985) Aromatic–aromatic interaction: a mechanism of protein structure stabilization. Science 229:23–28. doi:10.1126/science.3892686
- Kumar R, Liu D, Zhang L (2008) Advances in proteinous biomaterials. J Biobased Mater Bioenergy 2:1–24. doi:10.1166/jbmb.2008.204
- Reddy M, Mohanty AK, Misra M (2010) Thermoplastics from soy protein: a review on processing, blends and composites. J Biobased Mater Bioenergy 4:298–316. doi:10.1166/jbmb.2010.1112
- Sanger F, Thompson EO, Kitai R (1955) The amide groups of insulin. Biochem J 59:509–518
- Hagan ND, Higgins TJV (2004) Storage proteins and their metabolism. Handbook of plant biotechnology. Wiley, New York. doi:10.1002/0470869143.kc026
- Wool RP, Sun XS (2005) Bio-based polymers and composites. Elsevier/Academic Press, Amsterdam/Boston
- Krishna OD, Kiick KL (2010) Protein- and peptide-modified synthetic polymeric biomaterials. J Pept Sci 94:32–48. doi:10.1002/bip.21333
- Boyer C, Huang X, Whittaker MR, Bulmus V, Davis TP (2011) An overview of protein-polymer particles. Soft Matter 7:1599–1614. doi:10.1039/C0SM00412J
- Oh JH, McCurdy AR, Clark S, Swanson BG (2002) Characterization and thermal stability of polymorphic forms of synthesized tristearin. J Food Sci 67:2911–2917. doi:10.1111/j.1365-2621.2002.tb08837.x
- Cuq B, Gontard N, Guilbert S (1998) Proteins as agricultural polymers for packaging production. Cereal Chem 75:1–9. doi:10.1094/cchem.1998.75.1.1
- Noel TR, Parker R, Ring SG, Tatham AS (1995) The glass-transition behaviour of wheat gluten proteins. Int J Biol Macromol 17:81–85. doi:10.1016/0141-8130(95)93521-X
- Pedram Rad Z, Tavanai H, Moradi AR (2012) Production of feather keratin nanopowder through electrospraying. J Aerosol Sci 51:49–56. doi:10.1016/j.jaerosci.2012.04.007
- Eslahi N, Dadashian F, Nejad NH (2013) An investigation on keratin extraction from wool and feather waste by enzymatic hydrolysis. Prep Biochem Biotechnol 43:624–648. doi:10.1080/10826068.2013.763826
- Rao DR, Gupta VB (1992) Thermal-characteristics of wool fibers. J Macromol Sci Phys B31:149–162. doi:10.1080/0022349208215509
- Frikha M, Valencia DG, de Coca-Sinova A, Lazaro R, Mateos GG (2013) Ileal digestibility of amino acids of unheated and autoclaved pea protein concentrate in broilers. Poult Sci 92:1848–1857. doi:10.3382/ps.2013-03007
- Adebiyi AP, Aluko RE (2011) Functional properties of protein fractions obtained from commercial yellow field pea (*Pisum sativum* L.) seed protein isolate. Food Chem 128:902–908. doi:10.1016/j.foodchem.2011.03.116
- Manamperi WAR, Wiesenborn DP, Chang SKC, Pryor SW (2011) Effects of protein separation conditions on the functional and thermal properties of canola protein isolates. J Food Sci 76:E266–E273. doi:10.1111/j.1750-3841.2011.02087.x
- Wu J, Muir AD (2008) Comparative structural, emulsifying, and biological properties of 2 major canola proteins, cruciferin and napin. J Food Sci 73:c210; IX–IX
- Owusu-Ansah YJ, McCurdy SM (1991) Pea proteins: a review of chemistry, technology of production, and utilization. Food Rev Int 7:103–134
- Ullah A, Vasanthan T, Bressler D, Elia AL, Wu JP (2011) Bioplastics from feather quill. Biomacromolecules 12:3826–3832. doi:10.1021/bm201112n
- Uruakpa FO, Arntfield SD (2006) Surface hydrophobicity of commercial canola proteins mixed with kappa-carrageenan or guar gum. Food Chem 95:255–263. doi:10.1016/j.foodchem.2005.01.030
- Cabra V, Arreguin R, Vazquez-Duhalt R, Farres A (2007) Effect of alkaline deamidation on the structure, surface hydrophobicity, and emulsifying properties of the z19 alpha-zein. J Agric Food Chem 55:439–445. doi:10.1021/jf061002r
- Liang H-N, Tang CH (2013) Ph-dependent emulsifying properties of pea [*Pisum sativum* (L.)] proteins. Food Hydrocoll 33:309–319. doi:10.1016/j.foodhyd.2013.04.005
- Gharsallaoui A, Cases E, Chambin O, Saurel R (2009) Interfacial and emulsifying characteristics of acid-treated pea protein. Food Biophys 4:273–280. doi:10.1007/s11483-009-9125-8
- Toni M, Alibardi L (2007) Soft epidermis of a scaleless snake lacks beta-keratin. Eur J Histochem 51:145–151
- Cabra V, Arreguin R, Galvez A, Quirasco M, Vazquez-duhalt R, Farres A (2005) Characterization of a 19 kDa  $\alpha$ -zein of high purity. J Agric Food Chem 53:725–729. doi:10.1021/jf048530s
- Zhang J, Li Y, Li JS et al (2013) Isolation and characterization of biofunctional keratin particles extracted from wool wastes. Powder Technol 246:356–362. doi:10.1016/j.powtec.2013.05.037
- Lagrain B, Goderis B, Brijs K, Delcour JA (2010) Molecular basis of processing wheat gluten toward biobased materials. Biomacromolecules 11:533–541. doi:10.1021/bm100008p
- Park HJ, Bunn JM, Weller CL, Vergano PJ, Testin RF (1994) Water-vapor permeability and mechanical-properties of grain protein-based films as affected by mixtures of polyethylene-glycol and glycerin plasticizers. Trans ASAE 37:1281–1285
- Gennadios A, Weller CL, Testin RF (1993) Property modification of edible wheat, gluten-based films. Trans ASAE 36:465–470
- Donhowe IG, Fennema O (1993) The effects of plasticizers on crystallinity, permeability, and mechanical properties of

- methylcellulose films. *J Food Process Preserv* 17:247–257. doi:10.1111/j.1745-4549.1993.tb00729.x
38. Lieberman ER, Gilbert SG (1973) Gas permeation of collagen films as affected by cross-linkage, moisture, and plasticizer content. *J Polym Sci* 41:33–43. doi:10.1002/polc.5070410106
  39. Cuq B, Gontard N, Cuq JL, Guilbert S (1997) Selected functional properties of fish myofibrillar protein-based films as affected by hydrophilic plasticizers. *J Agric Food Chem* 45:622–626. doi:10.1021/jf960352i
  40. McHugh TH, Krochta JM (1994) Sorbitol- vs glycerol-plasticized whey protein edible films: integrated oxygen permeability and tensile property evaluation. *J Agric Food Chem* 42:841–845. doi:10.1021/jf00040a001
  41. Gontard N, Duchez C, Cuq JL, Guilbert S (1994) Edible composite films of wheat gluten and lipids: water vapour permeability and other physical properties. *Int J Food Sci Technol* 29:39–50. doi:10.1111/j.1365-2621.1994.tb02045.x
  42. Vieira MGA, da Silva MA, dos Santos LO, Beppu MM (2011) Natural-based plasticizers and biopolymer films: a review. *Eur Polym J* 47:254–263. doi:10.1016/j.eurpolymj.2010.12.011
  43. Rouilly A, Rigal L (2002) Agro-materials: a bibliographic review. *J Macromol Sci C42*:441–479. doi:10.1081/mc-120015987
  44. Kumar R, Choudhary V, Mishra S, Varma IK, Mattiason B (2002) Adhesives and plastics based on soy protein products. *Ind Crops Prod* 16:155–172. doi:10.1016/S0926-6690(02)00007-9
  45. Kalapathy U, Hettiarachchy NS, Rhee KC (1997) Effect of drying methods on molecular properties and functionalities of disulfide bond-cleaved soy proteins. *J Am Oil Chem Soc* 74:195–199. doi:10.1007/s11746-997-0123-z
  46. Kawamura Y, Matsumura Y, Matoba T, Yonezawa D, Kito M (1985) Selective reduction of interpolypeptide and intrapolypeptide disulfide bonds of wheat glutenin from defatted flour. *Cereal Chem* 62:279–283
  47. Kim SH, Kinsella JE (1986) Effects of reduction with dithiothreitol on some molecular properties of soy glycinin. *J Agric Food Chem* 34:623–627. doi:10.1021/jf00070a009
  48. Geno technology Inc (2013) Protein modification agents. <http://www.gbiosciences.com/ResearchUploads/TRDocumentsFile>. Accessed 02 Mar 2013
  49. Domb AJ, Kost J, Wiseman D (1998) Handbook of biodegradable polymers. Taylor & Francis, Amsterdam
  50. Micard V, Morel MH, Bonicel J, Guilbert S (2001) Thermal properties of raw and processed wheat gluten in relation with protein aggregation. *Polymer* 42:477–485. doi:10.1016/S0032-3861(00)00358-X
  51. Ullsten NH, Gällstedt M, Johansson E, Gräslund A, Hedenqvist MS (2006) Enlarged processing window of plasticized wheat gluten using salicylic acid. *Biomacromolecules* 7:771–776. doi:10.1021/bm050822u
  52. Wu QX, Zhang LN (2001) Properties and structure of soy protein isolate-ethylene glycol sheets obtained by compression molding. *Ind Eng Chem Res* 40:1879–1883. doi:10.1021/ie000694k
  53. Swain SN, Biswal SM, Nanda PK, Nayak P (2004) Biodegradable soy-based plastics: opportunities and challenges. *J Polym Environ* 12:35–42. doi:10.1023/B:JOEE.0000003126.14448.04
  54. Gonzalez-Gutierrez J, Partal P, Garcia-Morales M, Gallegos C (2011) Effect of processing on the viscoelastic, tensile and optical properties of albumen/starch-based bioplastics. *Carbohydr Chem* 84:308–315. doi:10.1016/j.carbpol.2010.11.040
  55. Zarate-Ramirez LS, Martinez I, Romero A, Partal P, Guerrero A (2011) Wheat gluten-based materials plasticised with glycerol and water by thermoplastic mixing and thermomoulding. *J Sci Food Agric* 91:625–633. doi:10.1002/jsfa.4224
  56. Cherian G, Gennadios A, Weller C, Chinachoti P (1995) Thermomechanical behavior of wheat gluten films—effect of sucrose, glycerin, and sorbitol. *Cereal Chem* 72:1–6
  57. Slade L, Levine H, Ievolella J, Wang M (1993) The glassy state phenomenon in applications for the food industry: application of the food polymer science approach to structure–function relationships of sucrose in cookie and cracker systems. *J Sci Food Agric* 63:133–176. doi:10.1002/jsfa.2740630202
  58. Food and Agriculture Organization of the United Nations (2010) Maize, rice and wheat: area harvested, production quantity, yield. <http://faostat3.fao.org/faostat-gateway/go/to/download/Q/QC/E>
  59. Anderson TJ, Lamsal BP (2011) Zein extraction from corn, corn products, and coproducts and modifications for various applications: a review. *Cereal Chem* 88:159–173. doi:10.1094/cchem-06-10-0091
  60. Shukla R, Cheryan M (2001) Zein: the industrial protein from corn. *Ind Crop Prod* 13:171–192. doi:10.1016/S0926-6690(00)00064-9
  61. Brauer S, Meister F, Gottlober RP, Nechwatal A (2007) Preparation and thermoplastic processing of modified plant proteins. *Macromol Mater Eng* 292:176–183. doi:10.1002/mame.200600364
  62. Hoseney RC (1986) Principles of cereal science and technology. American Association of Cereal Chemists, St. Paul
  63. Wilson CM (1988) Electrophoretic analyses of various commercial and laboratory-prepared zeins. *Cereal Chem* 65:72–73
  64. Lawton JW (2002) Zein: a history of processing and use. *Cereal Chem* 79:1–18. doi:10.1094/cchem.2002.79.1.1
  65. Lawton JW (2004) Plasticizers for zein: their effect on tensile properties and water absorption of zein films. *Cereal Chem* 81:1–5. doi:10.1094/cchem.2004.81.1.1
  66. Shi K, Yu H, Lakshmana Rao S, Lee TC (2012) Improved mechanical property and water resistance of zein films by plasticization with tributyl citrate. *J Agric Food Chem* 60:5988–5993. doi:10.1021/jf3001444
  67. Gong S, Wang H, Sun Q, Xue ST, Wang JY (2006) Mechanical properties and in vitro biocompatibility of porous zein scaffolds. *Biomaterials* 27:3793–3799. doi:10.1016/j.biomaterials.2006.02.019
  68. Xu H, Chai Y, Zhang G (2012) Synergistic effect of oleic acid and glycerol on zein film plasticization. *J Agric Food Chem* 60:10075–10081. doi:10.1021/jf302940j
  69. Gillgren T, Barker SA, Belton PS, Georget DMR, Stading M (2009) Plasticization of zein: a thermomechanical, ftir, and dielectric study. *Biomacromolecules* 10:1135–1139. doi:10.1021/bm801374q
  70. Gao C, Stading M, Wellner N et al (2006) Plasticization of a protein-based film by glycerol: a spectroscopic, mechanical, and thermal study. *J Agric Food Chem* 54:4611–4616. doi:10.1021/jf060611w
  71. Parris N, Coffin DR (1997) Composition factors affecting the water vapor permeability and tensile properties of hydrophilic zein films. *J Agric Food Chem* 45:1596–1599. doi:10.1021/jf960809o
  72. Lai HM, Padua GW, Wei LS (1997) Properties and microstructure of zein sheets plasticized with palmitic and stearic acids. *Cereal Chem* 74:83–90. doi:10.1094/cchem.1997.74.1.83
  73. Wang Q, Padua GW (2005) Properties of zein films coated with drying oils. *J Agric Food Chem* 53:3444–3448. doi:10.1021/jf047994n
  74. Zhang ML, Reitmeier CA, Hammond EG, Myers DJ (1997) Production of textile fibers from zein and a soy protein zein blend. *Cereal Chem* 74:594–598. doi:10.1094/cchem.1997.74.5.594
  75. Yang YQ, Wang LM, Li SQ (1996) Formaldehyde-free zein fiber—preparation and investigation. *J Appl Polym Sci*



- 59:433–441. doi:10.1002/(sici)1097-4628(19960118)59:3<433:aid-app7>3.0.co;2-q
76. Yamada K, Takahashi H, Noguchi A (1995) Improved water resistance in edible zein films and composites for biodegradable food packaging. *Int J Food Sci Technol* 30:599–608
  77. Kim S, Sessa DJ, Lawton JW (2004) Characterization of zein modified with a mild cross-linking agent. *Ind Crop Prod* 20:291–300. doi:10.1016/j.indcrop.2003.10.013
  78. Shi K, Huang YP, Yu HL, Lee TC, Huang QR (2011) Reducing the brittleness of zein films through chemical modification. *J Agric Food Chem* 59:56–61. doi:10.1021/jf103164r
  79. Selling GW, Woods KK, Biswas A, Willett JL (2009) Reactive extrusion of zein with glyoxal. *J Appl Polym Sci* 113:1828–1835. doi:10.1002/app.30074
  80. Arcan I, Yemencioğlu A (2011) Incorporating phenolic compounds opens a new perspective to use zein films as flexible bioactive packaging materials. *Food Res Int* 44:550–556. doi:10.1016/j.foodres.2010.11.034
  81. Del Nobile MA, Conte A, Incoronato AL, Panza O (2008) Antimicrobial efficacy and release kinetics of thymol from zein films. *J Food Eng* 89:57–63. doi:10.1016/j.jfoodeng.2008.04.004
  82. Janes ME, Kooshesh S, Johnson MG (2002) Control of listeria monocytogenes on the surface of refrigerated, ready-to-eat chicken coated with edible zein film coatings containing nisin and/or calcium propionate. *J Food Sci* 67:2754–2757. doi:10.1111/j.1365-2621.2002.tb08810.x
  83. Ha TT, Padua GW (2001) Effect of extrusion processing on properties of zein-fatty acids sheets. *Trans ASAE* 44:1223–1228
  84. Kruger Woods K, Selling GW (2008) Melt reaction of zein with glyoxal to improve tensile strength and reduce solubility. *J Appl Polym Sci* 109:2375–2383. doi:10.1002/app.28334
  85. Wang Y, Padua GW (2003) Tensile properties of extruded zein sheets and extrusion blown films. *Macromol Mater Eng* 288:886–893. doi:10.1002/mame.200300069
  86. Selling GW, Sessa DJ (2007) Sample preparation and testing methods affect the physical properties and evaluation of plasticized zein. *Ind Crop Prod* 25:266–273. doi:10.1016/j.indcrop.2006.12.010
  87. Oliviero M, Di Maio E, Iannace S (2010) Effect of molecular structure on film blowing ability of thermoplastic zein. *J Appl Polym Sci* 115:277–287. doi:10.1002/app.31116
  88. Santosa FXB, Padua GW (1999) Tensile properties and water absorption of zein sheets plasticized with oleic and linoleic acids. *J Agric Food Chem* 47:2070–2074
  89. Wang Y, Padua GW (2004) Water sorption properties of extruded zein films. *J Agric Food Chem* 52:3100–3105. doi:10.1021/jf035329t
  90. Woods KK, Selling GW (2007) Improved tensile strength of zein films using glyoxal as a crosslinking reagent. *J Biobased Mater Bioenergy* 1:282–288. doi:10.1166/jbmb.2007.033
  91. Ghanbarzadeh B, Musavi M, Oromiehie AR, Rezayi K, Rad ER, Milani J (2007) Effect of plasticizing sugars on water vapor permeability, surface energy and microstructure properties of zein films. *Lwt-Food Sci Technol* 40:1191–1197. doi:10.1016/j.lwt.2006.07.008
  92. Ghanbarzadeh B, Oromiehie AR, Musavi M, D-Jomeh ZE, Rad ER, Milani W (2006) Effect of plasticizing sugars on rheological and thermal properties of zein resins and mechanical properties of zein films. *Food Res Int* 39:882–890. doi:10.1016/j.foodres.2006.05.011
  93. Wang JY, Fujimoto K, Miyazawa T, Endo Y (1991) Antioxidative mechanism of maize zein in powder model systems against methyl linoleate—effect of water activity and coexistence of antioxidants. *J Agric Food Chem* 39:351–355. doi:10.1021/jf00002a025
  94. Wang JY, Miyazawa T, Fujimoto K (1991) Inhibition of methyl linoleate peroxidation by maize zein in powder model system at high water activity. *Agric Biochem* 55:1531–1536
  95. Herald TJ, Hachmeister KA, Huang S, Bowers JR (1996) Corn zein packaging materials for cooked turkey. *J Food Sci* 61:415–418. doi:10.1111/j.1365-2621.1996.tb14206.x
  96. Rakotonirainy AM, Wang Q, Padua GW (2001) Evaluation of zein films as modified atmosphere packaging for fresh broccoli. *J Food Sci* 66:1108–1111. doi:10.1111/j.1365-2621.2001.tb16089.x
  97. Dong J, Sun Q, Wang JY (2004) Basic study of corn protein, zein, as a biomaterial in tissue engineering, surface morphology and biocompatibility. *Biomaterials* 25:4691–4697. doi:10.1016/j.biomaterials.2003.10.084
  98. Edith Mathiowitz B, Howard Bernstein C, Eric Morrel N, Kristen Schwaller D (1993) Method for producing protein microspheres. US Patent 5,271,961, 21 Dec 1993
  99. Suzuki T, Sato E, Matsuda Y, Tada H, Unno K, Kato T (1989) Preparation of zein microspheres conjugated with antitumor drugs available for selective cancer-chemotherapy and development of a simple colorimetric determination of drugs in microspheres. *Chem Pharm Bull* 37:1051–1054
  100. Virgallito TT, Windschauer RJ (2012) Edible film useful for masking bitter taste, comprises a composition effective to mask the taste of semen. US Patent 13,455,520, 20 Dec 2012
  101. Hickman CP, Hickman CP, Hickman FM (1974) Integrated principles of zoology. C. V. Mosby, Saint Louis
  102. Bonser RHC, Dawson C (1999) The structural mechanical properties of down feathers and biomimicking natural insulation materials. *J Mater Sci Lett* 18:1769–1770. doi:10.1023/A:1006631328233
  103. Wrześniewska-Tosik K, Adamiec J (2007) Biocomposites with a content of keratin from chicken feathers. *Fibres Text East Eur* 15:106–112
  104. Papadopoulou MC (1985) Processed chicken feathers as feed-stuff for poultry and swine. A review. *Agric Wastes* 14:275–290. doi:10.1016/S0141-4607(85)80009-3
  105. Martelli SM, Laurindo JB (2012) Chicken feather keratin films plasticized with polyethylene glycol. *Int J Polym Mater* 61:17–29. doi:10.1080/00914037.2011.557809
  106. Barone JR, Schmidt WF, Liebner CFE (2005) Thermally processed keratin films. *J Appl Polym Sci* 97:1644–1651. doi:10.1002/app.21901
  107. Poole AJ, Church JS, Huson MG (2008) Environmentally sustainable fibers from regenerated protein. *Biomacromolecules* 10:1–8. doi:10.1021/bm8010648
  108. Schrooyen PMM, Dijkstra PJ, Oberthur RC, Bantjes A, Feijen J (2000) Partially carboxymethylated feather keratins. 1. Properties in aqueous systems. *J Agric Food Chem* 48:4326–4334. doi:10.1021/jf9913155
  109. Nagai Y, Nishikawa T (1970) Alkali solubilization of chicken feather keratin. *Agric Biol Chem* 34:16–22
  110. Dalev PG (1994) Utilization of waste feathers from poultry slaughter for production of a protein-concentrate. *Bioresour Technol* 48:265–267. doi:10.1016/0960-8524(94)90156-2
  111. Ghorpade VM, Li H, Gennadios A, Hanna MA (1995) Chemically modified soy protein films. *Trans ASAE* 38:1805–1808
  112. Schrooyen PMM, Dijkstra PJ, Oberthur RC, Bantjes A, Feijen J (2000) Partially carboxymethylated feather keratins. 2. Thermal and mechanical properties of films. *J Agric Food Chem* 49:221–230. doi:10.1021/jf0004154
  113. Yamauchi K, Yamauchi A, Kusonoki T, Kohda A, Konishi Y (1996) Preparation of stable aqueous solution of keratins, and physicochemical and biodegradational properties of films. *J Biomed Mater Res* 31:439–444. doi:10.1002/(SICI)1097-4636(199608)31:4<439:AID-JBM1>3.0.CO;2-M



114. Maria Martelli S, Moore G, Silva Paes S, Gandolfo C, Laurindo JB (2006) Influence of plasticizers on the water sorption isotherms and water vapor permeability of chicken feather keratin films. *Lwt-Food Sci Technol* 39:292–301. doi:10.1016/j.lwt.2004.12.014
115. Poole A, Lyons R, Church J (2011) Dissolving feather keratin using sodium sulfide for bio-polymer applications. *J Polym Environ* 19:995–1004. doi:10.1007/s10924-011-0365-6
116. Rocha Plácido Moore G, Maria Martelli S, Gandolfo C, José do Amaral Sobral P, Borges Laurindo J (2006) Influence of the glycerol concentration on some physical properties of feather keratin films. *Food Hydrocoll* 20:975–982. doi:10.1016/j.foodhyd.2005.11.001
117. Ratnayake WS, Hoover R, Warkentin T (2002) Pea starch: composition, structure and properties—a review. *Starch Stärke* 54:217–234. doi:10.1002/1521-379X(200206)54:6<217:AID-STAR217>3.0.CO;2-R
118. Barone JR, Arikian O (2007) Composting and biodegradation of thermally processed feather keratin polymer. *Polym Degrad Stab* 92:859–867. doi:10.1016/j.polymdegradstab.2007.01.030
119. Sharma S, Hodges JN, Luzinov I (2008) Biodegradable plastics from animal protein coproducts: feathermeal. *J Appl Polym Sci* 110:459–467. doi:10.1002/app.28601
120. Bietz JA, Lookhart GL (1996) Properties and non-food potential of gluten. *Cereal Food World* 41:376–382
121. Gennadios A, Weller CL (1990) Edible films and coatings from wheat and corn proteins. *Food Technol Chic* 44:63–69
122. Tanabe T, Okitsu N, Tachibana A, Yamauchi K (2002) Preparation and characterization of keratin–chitosan composite film. *Biomaterials* 23:817–825. doi:10.1016/S0142-9612(01)00187-9
123. Yamauchi K, Maniwa M, Mori T (1998) Cultivation of fibroblast cells on keratin-coated substrata. *J Biomater Sci Polym Ed* 9:259–270. doi:10.1163/156856298x00640
124. Katoh K, Shibayama M, Tanabe T, Yamauchi K (2004) Preparation and physicochemical properties of compression-molded keratin films. *Biomaterials* 25:2265–2272. doi:10.1016/j.biomaterials.2003.09.021
125. Saskatchewan Ministry of Agriculture (2012) Dry pea fact sheet. <http://www.agriculture.gov.sk.ca/Default.aspx?DN=0c18f233-c517-4510-b398-e08fd216aad2>. Accessed 12 Jan 2013
126. Sosulski FW, Imafidon GI (1990) Amino acid composition and nitrogen-to-protein conversion factors for animal and plant foods. *J Agric Food Chem* 38:1351–1356
127. Salunkhe DK, Kadam SS (1989) *Crc handbook of world food legumes: nutritional chemistry, processing technology, and utilization*. CRC Press, Boca Raton
128. Schroeder HE (1982) Quantitative studies on the cotyledonary proteins in the genus *Pisum*. *J Sci Food Agric* 33:623–633
129. Casey R, Wrigley CW (1982) The importance of seed proteins in human-nutrition. *Qual Plant* 31:189–190. doi:10.1007/bf01108629
130. Boulter D (1983) Protein-composition of grains of the leguminosae. *Qual Plant* 32:247–252. doi:10.1007/bf01091189
131. Leterme P, Monmart T, Baudart E (1990) Amino acid composition of pea (*Pisum sativum*) proteins and protein profile of pea flour. *J Sci Food Agric* 53:107–110. doi:10.1002/jsfa.2740530112
132. Maurice TJ, Murray ED, Grealy JA (1982) Simulated adipose tissue and process for its production. US Patent 4,388,333, 7 Feb 1982
133. Fredrikson M, Biot P, Alminger ML, Carlsson NG, Sandberg AS (2001) Production process for high-quality pea-protein isolate with low content of oligosaccharides and phytate. *J Agric Food Chem* 49:1208–1212. doi:10.1021/jf000708x
134. De Graaf LA, Harmsen PFH, Vereijken JM, Mönikes M (2001) Requirements for non-food applications of pea proteins a review. *Nahrung* 45:408–411. doi:10.1002/1521-3803(20011001)45:6<408::AID-FOOD408>3.0.CO;2-#
135. Choi WS, Han JH (2002) Film-forming mechanism and heat denaturation effects on the physical and chemical properties of pea-protein-isolate edible films. *J Food Sci* 67:1399–1406. doi:10.1111/j.1365-2621.2002.tb10297.x
136. Choi W-S, Han JH (2001) Physical and mechanical properties of pea-protein-based edible films. *J Food Sci* 66:319–322. doi:10.1111/j.1365-2621.2001.tb11339.x
137. Gueguen J, Viroben G, Noireaux P, Subirade M (1998) Influence of plasticizers and treatments on the properties of films from pea proteins. *Ind Crop Prod* 7:149–157. doi:10.1016/S0926-6690(97)00043-5
138. Viroben G, Barbot J, Mouloungui Z, Guéguen J (2000) Preparation and characterization of films from pea protein. *J Agric Food Chem* 48:1064–1069. doi:10.1021/jf9813891
139. Kowalczyk D, Baraniak B (2011) Effects of plasticizers, pH and heating of film-forming solution on the properties of pea protein isolate films. *J Food Eng* 105:295–305. doi:10.1016/j.jfoodeng.2011.02.037
140. Gunstone FD (2004) *Rapeseed and canola oil: production, processing, properties and uses*. CRC Press, Oxford
141. Shahidi F (1990) *Canola and rapeseed: production, chemistry, nutrition, and processing technology*. Van Nostrand Reinhold, New York
142. Tan SH, Mailer RJ, Blanchard CL, Agboola SO (2011) Canola proteins for human consumption: extraction, profile, and functional properties. *J Food Sci* 76:R16–R28. doi:10.1111/j.1750-3841.2010.01930.x
143. Arai KM, Takahashi R, Yokote Y, Akahane K (1983) Amino acid sequence of feather keratin from fowl. *Eur J Biochem* 132:501–507. doi:10.1111/j.1432-1033.1983.tb07389.x
144. Kimber DS, McGregor DI (1995) *Brassica oilseeds: production and utilization*. CAB International, Wallingford
145. Newkirk R, Canola Council of Canada (2009) *Canola meal feed industry guide 4th ed.* [http://www.canolacouncil.org/media/516716/canola\\_meal\\_feed\\_guide\\_english.pdf](http://www.canolacouncil.org/media/516716/canola_meal_feed_guide_english.pdf). Accessed 25 Nov 2013
146. Hoeglund AS, Roedin J, Larsson E, Rask L (1992) Distribution of napin and cruciferin in developing rape seed embryos. *J Plant Physiol* 98:509–515
147. Aider M, Barbana C (2011) Canola proteins: composition, extraction, functional properties, bioactivity, applications as a food ingredient and allergenicity—a practical and critical review. *Trends Food Sci Technol* 22:21–39. doi:10.1016/j.tifs.2010.11.002
148. Ghodsvali A, Khodaparast MHH, Vosoughi M, Diosady LL (2005) Preparation of canola protein materials using membrane technology and evaluation of meals functional properties. *Food Res Int* 38:223–231. doi:10.1016/j.foodres.2004.10.007
149. Tzeng YM, Diosady LL, Rubin LJ (1988) Preparation of rapeseed protein isolate by sodium hexametaphosphate extraction, ultrafiltration, diafiltration, and ion-exchange. *J Food Sci* 53:1537–1541. doi:10.1111/j.1365-2621.1988.tb09318.x
150. Tzeng YM, Diosady LL, Rubin LJ (1990) Production of canola protein materials by alkaline extraction, precipitation, and membrane processing. *J Food Sci* 55:1147. doi:10.1111/j.1365-2621.1990.tb01619.x
151. Manamperi WAR, Chang SKC, Ulven CA, Pryor SW (2010) Plastics from an improved canola protein isolate: preparation and properties. *J Am Oil Chem Soc* 87:909–915. doi:10.1007/s11746-010-1616-8
152. Pastuszewska B, Jablecki G, Swiech E, Buraczewska L, Ochtabinska A (2000) Nutritional value of rapeseed meal containing lecithin gums precipitated with citric acid. *Anim Feed Sci Technol* 86:117–123. doi:10.1016/s0377-8401(00)00162-0

153. Klockeman DM, Toledo R, Sims KA (1997) Isolation and characterization of defatted canola meal protein. *J Agric Food Chem* 45:3867–3870. doi:10.1021/jf970026i
154. Lasztity R, Hidvegi M (1985) Amino acid composition and biological value of cereal proteins. Kluwar Academic Publisher, Budapest
155. Hughes GJ, Ryan DJ, Mukherjea R, Schasteen CS (2011) Protein digestibility-corrected amino acid scores (pdcaas) for soy protein isolates and concentrate: criteria for evaluation. *J Agric Food Chem* 59:12707–12712
156. Goering KJ, Thomas OO, Beardsley DR, Curran WA (1960) Nutritional value of mustard and rape seed meals as protein source for rats. *J Nutr* 72:210–216
157. Niewiadomski H (1990) Rapeseed: chemistry and technology. Elsevier/PWN–Polish Scientific Publishers, Amsterdam/Warszawa
158. Shin YJ, Jang SA, Song HY, Song HJ, Bin Song K (2011) Effects of combined fumaric acid-uv-c treatment and rapeseed protein-gelatin film packaging on the postharvest quality of ‘seolhyang’ strawberries. *Food Sci Biotechnol* 20:1161–1165. doi:10.1007/s10068-011-0159-6
159. Li NB, Qi GY, Sun XS, Wang DH (2012) Effects of sodium bisulfite on the physicochemical and adhesion properties of canola protein fractions. *J Polym Environ* 20:905–915. doi:10.1007/s10924-012-0490-x
160. Thompson LU, Serrano MR (1986) Effect of phytic acid reduction on rapeseed protein digestibility and amino-acid-absorption. *J Agric Food Chem* 34:468–469. doi:10.1021/jf00069a023
161. Kishore NV, Murthy K, Rao MSN (1984) Interaction of phytate with mustard 12s protein. *J Agric Food Chem* 32:493–498
162. Ismond MAH, Welsh WD (1992) Application of new methodology to canola protein-isolation. *Food Chem* 45:125–127. doi:10.1016/0308-8146(92)90022-t
163. Jang SA, Lim GO, Song KB (2011) Preparation and mechanical properties of edible rapeseed protein films. *J Food Sci* 76:C218–C223. doi:10.1111/j.1750-3841.2010.02026.x
164. Wäsche A, Wurst S, Borcherdig A, Luck T (1998) Film forming properties of rape-seed protein after structural modification. *Nahrung* 42:269–271. doi:10.1002/(SICI)1521-3803(199808)42:03/04<269:AID-FOOD269>3.0.CO;2-B
165. Li B, Kennedy JF, Jiang QG, Xie BJ (2006) Quick dissolvable, edible and heatsealable blend films based on konjac glucomannan—gelatin. *Food Res Int* 39:544–549. doi:10.1016/j.foodres.2005.10.015
166. Krause JP (2002) Comparison of the effect of acylation and phosphorylation on surface pressure, surface potential and foaming properties of protein isolates from rapeseed (*Brassica napus*). *Ind Crop Prod* 15:221–228. doi:10.1016/s0926-6690(01)00117-0
167. Jang SA, Shin YJ, Song KB (2011) Effect of rapeseed protein-gelatin film containing grapefruit seed extract on ‘maehyang’ strawberry quality. *Int J Food Sci Technol* 46:620–625. doi:10.1111/j.1365-2621.2010.02530.x
168. Manamperi WAR, Fuqua MA, Ulven CA, Wiesenborn DP, Pryor SW (2011) Effect of canola protein extraction parameters on protein-based plastic properties. *J Biobased Mater Bioenergy* 5:500–506. doi:10.1166/jbmb.2011.1174
169. Manamperi WAR, Pryor SW (2012) Properties of canola protein-based plastics and protein isolates modified using sds and sdb. *J Am Oil Chem Soc* 89:541–549. doi:10.1007/s11746-011-1935-4
170. Shi B, Wang JH, Shi B, Wang JH (2009) Algae-blended compositions for thermoplastic articles. US Patent 12,430,968, 28 Apr 2009

# Developments in crops and management systems to improve lignocellulosic feedstock production

**Bruce Coulman and Ajay Dalai**, University of Saskatchewan, Saskatoon, Canada\*

**Emily Heaton**, Iowa State University, Ames, IA, USA

**Camilo Perez Lee and Mark Lefsrud**, McGill University, Ste-Anne-de-Bellevue, Quebec, Canada

**David Levin**, University of Manitoba, Winnipeg, Canada

**Peggy G. Lemaux**, University of California, Berkeley, CA, USA

**David Neale and Sharon P. Shoemaker**, University of California, Davis, CA, USA

**Jaswinder Singh, Donald L. Smith and Joann K. Whalen**, McGill University, Ste-Anne-de-Bellevue, Quebec, Canada

Received October 28, 2012; revised and accepted May 9, 2013

View online at Wiley Online Library ([wileyonlinelibrary.com](http://wileyonlinelibrary.com)); DOI: 10.1002/bbb.1418;  
*Biofuels, Bioprod. Bioref.* (2013)

**Abstract:** There is an urgent need to develop viable, renewable, sustainable energy systems that can reduce global dependence on fossil fuel sources of energy. Biofuels such as ethanol are being utilized as blends in surface transportation fuels and have the potential to improve sustainability and reduce greenhouse gas emissions in the short term. Bioethanol, the most widely used liquid biofuel, is currently produced by converting sugars or starches from feed crops into ethanol. Use of this fuel source displaces and draws water consumption away from agricultural crops, increases soil erosion by shifting land from perennial grasses to annual crops, and increases use of fertilizers and insecticides. In contrast, bioethanol made from lignocellulosic biomass feedstocks does not have these limitations and in addition, offers a larger resource base: the amount of cellulosic material available for potential use vastly outweighs the amount of available starch-based feedstock. Therefore, bioethanol from lignocellulosic biomass has attracted considerable interest from biofuel developers. This review is an update of some developments to optimize cellulose extraction from feedstock crops and to improve crop yields and logistics. It concludes that agricultural and forestry systems that incorporate lignocellulosic biomass crops can be designed for improved ecological function and energy use efficiency. Development of crops that have both desirable cell-wall traits and high biomass productivity under sustainable low-input conditions can significantly enhance the economics and efficiency of the conversion process. Optimizing the logistics of moving feedstock from field or forest to bio-refinery can significantly reduce costs of using lignocellulosic feedstocks. © 2013 Society of Chemical Industry and John Wiley & Sons, Ltd

**Keywords:** biofuels; lignocellulosic; feedstock; production system

Correspondence to: Donald L. Smith, Dept Plant Science, Faculty of Agricultural and Environmental Sciences, McGill University, 21111 Lakeshore Rd, Ste-Anne-de-Bellevue, Quebec H9X 3V9, Canada. E-mail: [Donald.smith@mcgill.ca](mailto:Donald.smith@mcgill.ca)

\*Note: Authors are given in alphabetical order



## Introduction

Four of the great challenges facing humanity during the twenty-first century will be energy supply, fresh water supply, climate change, and global food security, and each of these will be influenced by our choice of biofuels. Global energy demand is rising steeply in both developed and developing nations and, although new sources of oil such as the Bakken and Eagle Ford shales in the USA are currently easing American energy concerns, they will be short-lived. With each new oil discovery, the extraction technology becomes increasingly complicated, expensive, and fraught with environmental issues such as high water use and increased greenhouse gases (GHGs). The era of cheap oil is over. Climate change has the potential to substantially alter global habitats and crop-growing conditions, and reducing CO<sub>2</sub> emissions from energy use is a real and urgent necessity. Thus, we focus on biofuels derived from non-food biomass, and largely from lignocellulosic materials that can be used in solid (pelleted material, terrified biomass, etc.), liquid (ethanol or other alcohols, bio-oils, etc.) or gaseous (e.g. syngas) forms. Our first efforts at the production of biomass for biofuels involved feed grain crops but concerns about global food prices and supply have reoriented efforts toward developing biofuel crops that will grow on land not generally used for feed grains or to use plant-based residues not otherwise used. Currently, biofuels supply about 10% of the world's energy, with the bulk of this being low grade biomass used for cooking in developing countries. However, there are some examples of very successful biofuel development: Brazil currently produces about 50% of its liquid fuels on roughly 1% of its agricultural land and, through developing the resource, has positioned itself extremely well in a world where fossil fuels are relied on less and less. In the USA, total renewable fuels are forecast to be 16.55 billion gallons by 2013 (9.6% of total fuel) and advanced biofuels, including cellulosic, are forecast to be 2.75 billion gallons (1.6% of total fuel). Overall, the use of biofuels, especially the new forms that compete minimally with food production, is likely to rise, and the timeline of this transition suggests that a minimum of 10 million barrels per day of alternative fuels will be needed within a decade of the peak in production of conventional crude oil.<sup>4</sup> There is, therefore, an urgent need to develop viable, renewable, and sustainable energy systems that can displace global dependence on fossil fuel sources of energy. Although some sectors can manage with other energy sources, such as solar and wind-produced electricity, sectors such as the aerospace industry, are absolutely dependent on

energy-dense liquid fuels. Bioethanol, the most widely used liquid biofuel, can be utilized as a surface transportation fuel with little change to current technologies and has the potential to improve sustainability and reduce greenhouse gas emissions in the short term. Bioethanol is currently produced by first-generation technologies converting sugars directly from crops like sugarcane or sugarbeets, or indirectly through starch from corn, sorghum, or wheat. Domestic bioethanol production not only decreases dependence on fossil fuels, the addition of ethanol to gasoline also increases the fuel octane rating and results in cleaner, more complete combustion and lower GHG emissions. Combustion of 10% ethanol-blended gasoline results in production of 30% less carbon monoxide (CO), 10% less carbon dioxide (CO<sub>2</sub>), and 7% less NO<sub>x</sub>/SO<sub>x</sub>.<sup>5</sup>

## Generating biofuel from lignocellulosic biomass

First-generation bioethanol technologies, however, have several conspicuous limitations. Most notably is their total reliance on cultivated biomass and the diversion of feedgrains, such as corn to biofuel production.<sup>6</sup> In addition, there are environmental limitations including drawing water consumption away from land used to produce food crops, increased soil erosion through the shifting of land from perennial grasses to annual crops such as corn, and greater reliance on nitrogen and phosphate fertilizers, insecticides and herbicides.<sup>7–11</sup> The strongest argument against first-generation technologies, however, comes with the reality of their limited supply and competition with food. Even if all soybean and corn production in the USA were dedicated to biofuel production, only 12% of gasoline demand and 6% of diesel demand would be displaced.<sup>12</sup>

First-generation technologies are therefore not a solution to the world's long-term energy needs. Adopting present processing technologies to utilize a feedstock that does not require heavy cultivation and diversion of agricultural lands and foodstuffs could, however, contribute to a long-term solution to bioenergy generation and sustainable supply. Second-generation biofuels are made from lignocellulosic biomass feedstocks using advanced technological processes that convert cellulose, found in plant structural elements, to ethanol.<sup>13</sup> The amount of cellulosic material available for potential use vastly outweighs the amount of available starch-based substrate. A conservative estimate is that presently, there are approximately 400 million tons of biomass available and this number could grow to about 600 million by 2020.<sup>14</sup> The lignocellulosic feedstocks in Canada are cereal residues in agricultural regions and

AQ1



wood residues in forest regions. The annual availability of wood infected by pine beetle in Canada is between 9.3 and 12.3 million tonnes, which has the potential to generate 2.8–3.6 billion liters of ethanol per year.<sup>2</sup> Every year, more than 40 million tonnes of lignocellulosic biomass, that is produced worldwide, is thrown away. There is no competition with food when these materials are converted to biofuels.<sup>3</sup> In the longer term (by 2050) there is the potential worldwide to produce 130–410 EJ/year of energy, equivalent to 33 to 100% of present energy production, by using only abandoned agricultural lands, low-productivity lands and ‘rest lands’.<sup>15</sup> The cost of pre-processing cellulosic material to generate free glucose, however, is much higher than that for conventional feedstock, as both mechanical and thermochemical treatments are often required. However, conversion costs have been falling in recent years and pilot plants for the production of lignocellulosic fuels are now being developed. Lignocellulose biomass can also be used in the generation of heat and electricity through direct combustion.

Emerging demands for biofuels and bioenergy derived from biomass are creating new opportunities for redesigning agricultural and forestry systems for improved ecological function and efficient energy use. Technologies to optimize plant cell wall characteristics and reduce energy requirements for polysaccharide extraction, breeding, and agronomic management of feedstock production systems to increase productivity, and efficient harvesting techniques can all significantly enhance the economics and efficiency of second-generation biofuels and make them more cost effective as fossil fuel replacements. What follows is an update of some developments to optimize cellulose extraction from feedstock crops and to improve crop yields and harvesting technologies.

## Developments in cell wall degradation

Cellulose and hemicellulose, the main polysaccharides in lignocellulosic biomass are tightly bound to lignin in the plant cell wall, which hinders their availability for bioconversion to bioethanol.<sup>16</sup> Overcoming cell wall chemical and structural properties to extract the desired carbohydrates is currently an energy-intensive process. Reducing the energy requirements through the development of mechanisms for easier cell wall degradation is critical for the advancement of biofuel production from lignocellulosic biomass.

Separating lignin and cellulose currently requires heat and acid to remove the lignin and reducing or modifying

the initial lignin content of the biomass could partially replace this treatment.<sup>16,17</sup> In recent years, genetic modification of the lignin biosynthesis pathway has received a lot of attention.<sup>17</sup> Transgenic modification of this pathway to alter lignin composition in trees improved pulping efficiency.<sup>20</sup> It was observed that transforming poplar plants with antisense constructs of the lignin synthesis gene coding for 4-coumarate-CoA ligase (*Pt4CL*) resulted in significantly decreased (40%) lignin content.<sup>21</sup> However, negative effects of plant susceptibility to pathogens or harsh environmental conditions have been anticipated.<sup>22</sup> In crop plants, down-regulation of lignin synthesis genes improves saccharification efficiency, potentially eliminating the need for acid pre-treatment.<sup>23</sup> Although extensive research has been conducted in the area of genetic modification of lignin biosynthesis in trees and other dicots, it is still not clear how much one can extrapolate from dicots to grasses.<sup>18</sup> Another promising approach would be to engineer various cellulase/ligninase enzymes into crop biomass in order to deconstruct the biomass before bioprocessing and allow it to be more readily hydrolyzed to produce ethanol.<sup>18,19</sup> Various groups of cellulases, for example endoglucanases, exoglucanases and  $\beta$ -glucosidases, have been identified and successfully introduced into plants through genetic engineering to facilitate cellulose degradation. This approach resulted in the expression of bacterial endoglucanase E1 in model plant species, such as *Arabidopsis* and tobacco.<sup>24</sup> Specifically, a thermostable endo-1,4- $\beta$ -D-glucanase E1 from *Acidothermus cellulolyticus* was targeted to the apoplast of transgenic *Arabidopsis*.<sup>25</sup> The enzyme has a high temperature (81°C) optimum and activity is reduced at ambient temperatures. This suggests that the production of such enzymes in plants is possible by virtue of their limited enzymatic activity at temperatures compatible with plant growth. Based on these results, endoglucanase was introduced into maize and found to be active.<sup>26</sup> Conversion of rice and maize biomass to ethanol was improved by supplementing the process with thermostable endoglucanase expressed in transgenic rice.<sup>27</sup> Engineering *Festuca* with a fungal ferulic acid esterase targeted to the vacuole resulted in increased digestibility and reduced levels of cell wall esterified phenolics.<sup>28</sup> As heterologous expression of ligninase/endoglucanase is feasible in crop plants, transformation of feedstocks with similar thermostable cellulases will also be a useful first step in developing these crops for cellulosic ethanol production. There will be a need to ensure that these enzymes are not degraded by high temperatures during biomass pretreatment processes.



Despite such successes with engineering cellulose-degrading activities into crop plants, it is notable that besides cellulose, lignocellulosic biomass contains large amounts of hemicelluloses with high contents of five-carbon sugars. The most common structural polymer found in the hemicelluloses is a  $\beta$ -1,4-linked xylose polymer. These polymers are recalcitrant to hydrolysis by current technologies; however, hemicellulase enzymes introduced through genetic engineering can increase overall conversion of hemicellulose by xylanases in a synergistic fashion. Complete hydrolysis of hemicelluloses could lead to a dramatic improvement in the fermentative and extraction processes, which could further improve bioenergy potential of lignocellulosic biomass.<sup>16,29</sup> Recently, a synthetic, modified, codon optimized xylanase gene (*XynZ*) from *Clostridium thermocellum* was successfully expressed in transgenic tobacco plants.<sup>30</sup> Further increase in performance of these enzymes and improvement to resulting sugar yields are vital to improving the efficiency of the lignocellulosic biofuel/bioproduct industry.

Transgenic technologies can also play an important role in enhancing yield and stress tolerance in biofuel crops.<sup>18,31</sup> Gene expression can be targeted to the apoplast and vacuole through a specific signal peptide sequence such as *Pr1a*, from tobacco and potato proteinase inhibitors. Identifying factors that facilitate tolerance and survival during exposure to drought, freezing, and other abiotic and biotic stresses will be vital. Therefore, feedstocks such as perennial grasses and relevant tree species could be transformed for more effective weed, disease, and insect control.<sup>32</sup>

At the cellular level, a new generation of energy crops will be characterized by a cellulose and hemicellulose content that is more accessible and energy efficient to extract. These crops must also have high biomass production and produce an optimized amount of fuel per unit of biomass while maintaining crop production system sustainability with minimal water and fertilizer inputs.

## Perennial grass production systems

Interest in using perennial grass species as energy crops is fairly recent and relatively little breeding has therefore been done for this purpose. Breeding programs, conducted since 1936, focused on improvements for forage purposes such as better nutritive value which included higher digestibility, higher concentration of various minerals, and lower fiber characteristics which are not always useful for bioenergy. Improved forage yield has also been a common goal, which would be an advantage for bioenergy

production; however, progress has been slow, due to the outcrossing nature and genetic complexity of many of the species. Similarly, agronomic management of perennial grass species has traditionally centered on forage quality and productivity. In recent years, however, a concerted effort has been made to evaluate the genetic by environment interaction of leading potential biomass species in a range of field trials across the USA.<sup>33</sup> Yield gains realized through improved crop management for biomass have frequently been of the same magnitude as those targeted through breeding programs.<sup>34–36</sup> Beyond biomass, the traits targeted in improved grass populations for biofuel production will depend on the technology pathway used to convert biomass to fuel.<sup>37</sup> For example, lignin composition and concentration and cellulose and hemicellulose concentration are related to ethanol yield in a fermentation system.<sup>38</sup>

Warm season ( $C_4$ ) grass species, including switchgrass (*Panicum virgatum* L.), Miscanthus (*Miscanthus*  $\times$  *giganteus* Greef et Deu. and other species) and prairie cordgrass (*Spartina pectinata* Bosc ex Link.), have received the most attention for biomass improvement through breeding. Reviews of warm season grasses as biofuel feedstocks, including information on genetic improvement, have recently been published.<sup>37–41</sup> A number of cool season grasses have also received interest, including reed canarygrass (*Phalaris arundinacea* L.).<sup>42</sup> Below we highlight key developments in the breeding and agronomic management of leading warm season grasses used for biomass production.

## Switchgrass

Switchgrass is native to the prairie region of North America and has a number of characteristics which are desirable for use as a bioenergy feedstock, including high productivity, persistence, and wide adaptation.<sup>37,43</sup> Switchgrass has been evaluated for use as a bioenergy crop in the USA for more than 30 years and its history as the 'model' bioenergy crop by the US Department of Energy has been reviewed.<sup>44</sup> There are two distinct ecotypes of switchgrass, upland (mainly octaploid  $2n = 8x = 72$ ) and lowland (mainly tetraploid  $2n = 4x = 36$ ). The two ecotypes can be distinguished by cytoplasmic (chloroplast DNA) differences and nuclear DNA differences.<sup>45,46</sup>

Cultivars and existing native populations have not been found to be highly differentiated and are genetically diverse providing useful germplasm for selection for biomass yield.<sup>47</sup> Narrow sense heritabilities of biomass yield have been estimated to be low to medium.<sup>48,49</sup>

Yield components such as tiller density, tiller mass, and phytomer mass have been positively associated with biomass yield<sup>50,51</sup> and these components may have higher heritability than biomass yield. Development of hybrids to exploit heterosis has recently been investigated and a method for hybrid production proposed.<sup>52</sup> Tetraploid populations (largely lowland) were identified as different heterotic groups with F<sub>1</sub> hybrid populations showing 30–38% high parent heterosis.<sup>53</sup> F<sub>1</sub> hybrids of the lowland ‘Kanlow’ by upland ‘Summer’ are now in field trials across the

**AQ5** USA, in preparation for expected commercial release.<sup>191</sup> Genomic tool development is not as advanced in switchgrass as in some other crop species, but a transformation project involving down-regulation of lignin pathway genes in switchgrass has been initiated.<sup>54</sup>

Management of switchgrass for biomass is dependent on location and cultivar selection, with general divisions evident between upland and lowland ecotypes. Generally, lowland ecotypes are best suited for warmer, wetter growth environments and have a longer growing period than upland ecotypes.<sup>55</sup> They have different morphologies, with larger, taller and fewer stems being characteristic of lowland ecotypes, often leading to higher biomass yields.<sup>35,51</sup> Further, the ecotypes have been shown to differ in their susceptibility to foliar diseases, such as rust (*Puccinia* sp.), necessitating different pest protection strategies depending on cultivar selection.<sup>56–58</sup> Ecotypes are also differentially affected by herbicides,<sup>59</sup> harvest management,<sup>60</sup> and resource availability<sup>55</sup> leading to changes in both biomass yield and biomass composition.<sup>61,62</sup>

Numerous studies have evaluated nitrogen fertilizer application and harvest management in switchgrass, identifying these management practices as critical to not only crop productivity but also to the long-term stand persistence and GHG emission or sequestration. Within ecoregions, best management practices for nitrogen fertilizer and harvest time have been developed. In warmer, wetter climates it will likely be possible to harvest twice per year if nutrients are not limiting, but in temperate areas a single, late harvest leads to better fuel quality and more consistent yields.<sup>192</sup> A single harvest after senescence minimizes nitrogen removal in plant biomass, thus reducing the amount of fertilizer needed for optimal crop growth.<sup>63</sup>

**AQ6**

## Miscanthus

*Miscanthus* is a C<sub>4</sub> grass native to East Asia, for which the main commercial focus has been on a single, high yielding sterile hybrid, *Miscanthus* × *giganteus*, a spontaneous allotriploid hybrid of *M. sinensis* and *M. sacchariflorus*

with 57 chromosomes.<sup>64–65</sup> There is genetic diversity available from a large gene pool of different species within *Miscanthus* and related species and a high genotypic variation for cell wall composition among *Miscanthus* genotypes has been found.<sup>66</sup> There are also major differences in biomass yield and associated traits among species and among ecotypes within species.<sup>38</sup> Breeding goals include both the development of new sterile hybrids which outyield the existing *M.* × *giganteus* hybrid, as well as the development of adapted, high yielding seed-propagated cultivars of *M. sinensis* or *M. sacchariflorus*, which would reduce establishment costs,<sup>38</sup> although seeded varieties must be carefully evaluated for invasive potential.<sup>67,68</sup> Variation exists among and within species for first-year overwintering ability in northern Europe, with genotypes of *M. sinensis* and species hybrids being superior.<sup>69,70</sup> Low-density marker maps have been generated in *M. sinensis* and potential quantitative trait loci (QTL) identified for a number of traits associated with biomass production.<sup>71</sup> Chloroplast DNA marker loci containing single nucleotide polymorphisms (SNPs) were identified, which can be used to differentiate *Miscanthus* species and identify cytoplasmic gene pools.<sup>72</sup>

*Miscanthus* has only been grown for biomass in North America for a short time. The first published field trials of *Miscanthus* in the USA compared it to switchgrass in Illinois, and found it to yield at least twice as much biomass.<sup>73</sup> Recent years have seen an explosion of *Miscanthus* research in the USA, following decades of experience with it in Europe.<sup>74,75</sup> The comparative yield advantage of *M.* × *giganteus* increases as climates become cooler, making it likely the most productive bioenergy crop available for cool, temperate regions.<sup>76,77</sup> So far, the biggest barriers to *Miscanthus* adoption revolve around propagation, planting, and winter survival of the sterile clone *M.* × *giganteus*. Propagule (rhizome) costs are high, frontloading *M.* × *giganteus* production costs; however, if *M.* × *giganteus* rhizome costs in the USA drop to prices seen in Europe, it is cost-competitive with other biomass feedstocks, or cheaper.<sup>193</sup> The crop is especially sensitive during the first winter after planting and becomes increasing resilient as stands mature. Because vegetative planting material is so expensive, great care is typically given to establishing stands, with inputs similar to those seen in annual grain crops.<sup>78</sup> Following the establishment year, inputs are very low, giving it favourable economic and greenhouse gas budgets.<sup>79</sup> Like switchgrass, no clear recommendation exists for nitrogen fertility in *M.* × *giganteus* given its inconsistent response to N fertilizer.<sup>80</sup> Generally, nutrient removal is quite low if the crop is harvested after senescence.<sup>63</sup>

**AQ7**

## Prairie cordgrass

Prairie cordgrass is native to North America and is of interest for biomass production because of its high productivity and adaptation to marginal land that is subject to salinity and water fluctuations.<sup>81</sup> Breeding and genetic studies for this species are limited. Populations are either tetraploid ( $2n = 4x = 40$ ) or octoploid ( $2n = 8x = 80$ ) and recently a sterile hexaploid plant was identified.<sup>81</sup> High variation in biomass among seven natural populations in South Dakota, USA suggested that promising strains could be selected directly from these populations.<sup>45</sup>

Very little scientific research has been conducted on prairie cordgrass management for biomass. Research has shown that the crop does not easily produce or grow from seed, and vegetative propagation, similar to *M. × giganteus*, may prove a more reliable means of establishment.<sup>82</sup> Fertility and harvest requirements are not yet clear.

Management of perennial grasses for bioenergy will require the same thoughtful and logical development as was given to annual crops during the first Green Revolution. The heterogeneous nature of landscapes will require site-specific research and more management-intensive cropping practices, i.e., the inputs of resources such as fertilizer and fossil fuels may be reduced, but the need for informed, involved land managers will increase as we seek to harvest both fuel and food from crop land.

## Woody production systems

### Feedstock development

In the 2006 US Department of Energy's roadmap publication entitled *Breaking the Biological Barriers to Cellulosic Ethanol*, several forest trees species were identified as potential feedstock sources for cellulosic ethanol production. These included hybrid poplars, willows, silver maple, black locust, sycamore, sweetgum, and eucalyptus. In the Northern Hemisphere, poplar has been given the most attention and is the target of large breeding programs and plantations for solid wood and pulp and paper production. Hybrid poplars can be grown in many regions of the USA and Canada but to date the amount of land in industrial plantations is still quite limited. This could change if genetic improvement and/or conversion technologies develop sufficiently to produce a sustainable and economic feedstock. Three general approaches to feedstock improvement are being pursued: (i) Traditional hybrid breeding, (ii) Genomics-assisted breeding and (iii) Transgenic modification.<sup>83</sup> Traditional hybrid breeding involves making inter-specific crosses and selecting superior  $F_1$  individuals.

Some very successful hybrids include *Populus × canadensis* (*deltoides × nigra*), *P. × generosa* (*trichocarpa × deltoides*), *P. × tomentosa* (*alba × tremula*) and *P. × wettsteinii* (*tremula × tremuloides*). These hybrids are now being further improved through genomics-assisted breeding<sup>83</sup> and transgenic modification.<sup>84</sup> Genomics-assisted breeding involves making selections based on the genotype of an individual rather than its phenotype. In practice, a combined genotypic and phenotypic selection strategy is often employed. Discovery of genetic markers by phenotypic trait relationships is the first step and two general approaches have been used: (i) QTL mapping<sup>85</sup> and (ii) association mapping.<sup>86</sup> The earlier QTL approach lacked high-resolution mapping of markers to genes affecting phenotypes and did not lead to application in breeding programs. The association mapping approach has delivered high-resolution marker × trait relationships and is currently being implemented into breeding programs, including programs seeking to develop poplar hybrids as biofuel feedstocks. The final genetic improvement approach is to produce genetically modified varieties through transgenic technology. Genetic transformation technology was first developed in a few poplar varieties back in the 1980s using the *Agrobacterium* system and has been used to develop transgenic varieties of many kinds, including those with genes modified to produce altered wood chemistry properties.<sup>87</sup> The transgenic approach holds great potential but is currently limited in application in the USA and Canada due to regulatory restrictions on deployment of genetically modified (GM) trees. Furthermore, many growers desire FSC (Forest Stewardship Council) certification and would not deploy GM trees even if they were available.

### Production systems

Woody perennial production systems are expected to have positive effects on soil properties, biodiversity, energy balance, GHG mitigation, and carbon footprint compared to arable crops.<sup>88</sup> Fast-growing willows can be harvested in three to ten year cycles,<sup>89</sup> and hybrid poplars are cut from twelve to twenty-five years after planting,<sup>90</sup> although shorter rotation coppicing systems also exist. Both crops have excellent potential for simultaneous heat and power generation through burning of wood pellets/biomass, but are not yet good candidates for bioethanol production due to the challenge of efficiently converting woody feedstocks into liquid biofuel.

Short rotation plantations of woody feedstocks are often established on unimproved or abandoned farmland, due to the relative ease of clearing and cultivating land that was previously devoted to agriculture. The plantation



may be the only crop in the field, or it may be grown with an intercrop. Tree-based intercropping systems feature widely spaced tree rows (10 to 15 m apart) with annual crops growing between established tree rows. Such systems diversify the rural landscape and provide economic returns to producers whilst trees are becoming established. They are also expected to store more carbon than conventional cropping systems through two mechanisms: (i) by increasing carbon storage in the biomass of planted trees,<sup>91</sup> and (ii) by adding inputs of lignin-rich litter that is slowly decomposed and thus stabilized as soil organic carbon, which is consistent with the goals of soil carbon sequestration.<sup>93</sup> After 22 years, there was 12% more soil organic carbon in a tree-based intercropping system than an adjacent conventional agroecosystem in Guelph, Ontario (Canada); the annual crop rotation in both systems was a corn-soybean-cereal rotation.<sup>93</sup>

Woody feedstocks are well adapted for cool, temperate climates. Twelve clones of fast-growing trees established on abandoned farmland in southern Quebec, Canada, accumulated biomass at rates of 66 to 72 t dry matter ha<sup>-1</sup> (hybrid poplar) and 62 to 68 t dry matter ha<sup>-1</sup> (willow) after four growing seasons.<sup>94</sup> In plantations, these woody perennials grow optimally when given sufficient space to avoid interspecific competition (around 40 000 willow ha<sup>-1</sup> and 2000 hybrid poplars ha<sup>-1</sup>), adequate NPK fertilization, as well as weed and insect control.<sup>89,90,95</sup> Clone selection is critical, as wood density, fiber content, fiber length, and other feedstock characteristics important for heat and power generation are strongly controlled by genetic traits. The estimated clonal repeatability for wood density and fiber length were much greater than for growth traits (diameter at breast height, tree height).<sup>96</sup>

The response of woody perennials to climate change must be considered, due to the fact that some trees will grow for more than a decade before biomass is harvested. Under elevated CO<sub>2</sub> concentrations, there is significant increase in biomass accumulation and lignin deposition in hybrid poplar wood.<sup>97</sup> Yet, elevated CO<sub>2</sub> and N fertilization did not affect the calorific value of wood, which was 19.3 MJ kg<sup>-1</sup>. Judicious use of N fertilizer enhances the energy production per land area because the yields of woody biomass are enhanced by 50% or more, compared to trees grown on unfertilized land.<sup>97</sup> Future climate scenarios also suggest warmer, drier growing conditions, so the development of new genotypes with high water use efficiency and watershed-scale management plans that consider the land and water requirements of woody feedstocks is recommended.<sup>88</sup>

The development of dedicated woody feedstocks for cellulosic ethanol is still in its early days, although significant

research investments are being made in both the USA and Canada. Woody feedstocks may offer important benefits for the environment and contribute significantly as an alternative feedstock for energy production. Further work is needed to develop policies that effectively manage short rotation plantations in the context of climate change and consider the hydrological implications of including woody feedstocks at the landscape scale. The whole life cycle of woody feedstock production (clone and site selection, management and production decisions, harvesting, transport, and energy transformation) needs to be within a regulatory framework where sustainability is a central driver.

## Sustainable solutions for feedstock production systems

Feedstock production systems must demonstrate positive net energy balances and be able to grow on land that is marginal for food production. In addition, they must use minimal amounts of water and where possible, increase soil organic matter levels and stabilize soils against erosion. Reducing energy consumption through use of conservation tillage rather than conventional tillage and utilizing crops, possibly as part of a rotation, that have low water demands may be useful in some regions. For example, sorghum can be used as a bioenergy crop in arid and semi-arid lands as the water required for its production is much lower than switchgrass and *Miscanthus*.<sup>98,102</sup> AQ8 Rocateli<sup>103</sup> studied grain sorghum (GS), high biomass forage sorghum (FS), photoperiod-sensitive forage sorghum (PS) under conventional and conservation tillage conditions and found that PS under conservation tillage produced the highest biomass and was the recommended bioenergy crop. From the perspective of this review, the material remaining after the seeds were removed, as a food or feed material, would be of interest as a biofuel feedstock.

Other factors can also improve plant growth while reducing inputs and thereby contribute to the development of low input, sustainable lignocellulosic production systems.

## Plant-growth-promoting rhizobacteria (PGPR)

PGPR include bacteria in the soil near plant roots, on the surface of plant root systems, in spaces between root cells or inside specialized cells of root nodules.<sup>104</sup> PGPR increase plant growth through a broad range

of mechanisms such as production of phytohormones (directly stimulating aspects of plant development and growth) or metal-chelating siderophores (making plant nutrients, such as iron, more available) and disease suppression through antibiosis.<sup>105</sup> Although a number of PGPR mechanisms are now understood, there is still much to discover regarding how bacteria-plant associations affect plant growth. At a time when we are looking to plants to provide biofuels and other novel bioproducts,<sup>106</sup> while still feeding the world's growing population, understanding mechanisms that can serve to increase overall plant productivity is increasingly imperative.

### PGPR-to-plant signals compounds

Lipo-chito-oligosaccharides (LCOs), a group of N<sub>2</sub> fixing PGPR, can alter the course of growth and development in a range of plants.<sup>107–113</sup> Enhanced germination and seedling growth, along with the mitogenic nature of LCOs, suggest accelerated meristem activity. LCOs, isolated from *B. japonicum*, accelerated seed germination, seedling emergence, root growth and development in soybean and non-leguminous plants,<sup>107,108</sup> and these effects were greater when the plants were under some level of stress. LCOs stimulated root growth in *Medicago truncatula*,<sup>112</sup> accelerated flowering (a typical response to stress) and increased yield when sprayed on tomatoes.<sup>113</sup> Foliar application of LCOs also induced resistance of soybean plants to powdery mildew.<sup>114</sup> Given that LCOs induce defence responses in *Medicago* cell cultures and roots,<sup>115</sup> that LCOs show structural similarity to chitin (they have a chitin backbone), and that chitin induces defence responses in plants, it is reasonable to hypothesize that LCOs induce aspects of plant defence responses similar to chitin. These defense mechanisms can aid biofuel feedstock crops in resisting both biotic (pathogen) and abiotic (cold, drought, etc.) stresses, leading to greater yields.

Bai *et al.*<sup>116</sup> isolated a PGPR, *B. thuringiensis* NEB 17, from soybean root nodules and showed that it enhances soybean nodulation and N<sub>2</sub> fixation when co-inoculated with *B. japonicum*.<sup>117</sup> The liquid medium that was used to grow NEB17 for plant growth stimulating materials was shown to contain a 31 kDa peptide, now named Thuricin-17<sup>118,119</sup> which, when sprayed on leaves or applied to roots, stimulates growth of corn and soybean, in a manner similar to that caused by LCOs.<sup>120,121</sup> Thuricin-17 is not toxic to *B. japonicum* 532C.<sup>119</sup> Bacteriocins are bacteria-produced peptides that are either bactericidal or bacteriostatic to specific bacterial strains that compete most closely with the producer strains. Bacteriocins are often isolated from bacteria found in food, such as strains

of *Bacillus*.<sup>122</sup> However, some bacteriocins have been isolated from extracellular PGPR (ePGPR), such as *Bacillus thuringiensis* subsp. kurstaki,<sup>123</sup> *Pseudomonas* spp.<sup>124</sup> and the nodulating intracellular PGPR (iPGPR), such as *Rhizobium leguminosarum* 248.<sup>125</sup> It has been postulated that bacteriocins produced by PGPR provide a competitive advantage to the producer strains<sup>126</sup> and may enhance nodule occupancy when the producer strain is one of the rhizobia.<sup>125</sup> Clearly, there are some previously unknown mechanisms at play and these could be exploited in the development of low-input biofuel feedstock production systems.

### Nitrogen fixers

Micro-organisms capable of biological nitrogen fixation (BNF) are largely beneficial soil bacteria and include rhizobia and free-living diazotrophs. These N<sub>2</sub>-fixing bacteria are collectively considered to be PGPR and are often found near, on, or within plant roots.<sup>104,127,128</sup> The success of bioethanol production from sugarcane in Brazil, has been attributed to lower inputs of N fertilizer since up to 80% of the plant N is derived from biological N<sub>2</sub>-fixation by associated PGPR.<sup>129,130</sup> The diazotrophs isolated from sugarcane include *Azospirillum* and *Acetobacter* or *Gluconacetobacter* species, as well as endophytic diazotrophs of the genera *Herbaspirillum* and *Burkholderia*.<sup>131,132</sup> Members of the diazotrophic genus *Azospirillum* are important sources of N<sub>2</sub> fixation and N transfer to many plants.<sup>132</sup> *G. diazotrophicus*, the predominant diazotroph of sugarcane, has also been shown to colonize rice, wheat, maize, and *Arabidopsis thaliana*.<sup>133</sup> Inoculation of *Herbaspirillum seropedicae* onto rice seedlings increased N content by 30%,<sup>134</sup> while inoculation of *Azospirillum lipoferum* and *A. brasilense*, isolated from kallar grass, onto rice provided nearly 70% of fixed nitrogen.<sup>135</sup> Potential utilization of BNF in the growth of cellulosic feedstock crops would significantly reduce N fertilizer and thus energy requirements associated with their production.

### Mycorrhizae

Mycorrhizal fungi constitute a very ancient symbiosis between higher plants and fungi.<sup>136</sup> The relationship is so well developed that the fungi often cannot grow in the absence of the host plant.<sup>137</sup> The fungi improve the ability of plants to take up soil P and Zn by effectively increasing root surface area and their ability to take up low mobility nutrients such as P.<sup>138</sup> These fungi are present in almost all soils of the world but selection for enhanced types and effective inoculation strategies can improve crop yields.<sup>139</sup>



Phosphorus conservation is particularly important as peak extraction of this nutrient is forecast for as little as 20 years from now, which could place a strong limitation on biomass and food crop production systems. Thus, biofuel production systems and very effective mycorrhizal systems, able to make the most of this P, will be critical in the near future.

## Biochar

Biochar is black, carbon-rich material produced when organic matter is thermally cracked in an oxygen-limited or oxygen-free environment (pyrolysis). The particles of char produced this way contain primarily carbon and inorganic matter (ash), are highly porous<sup>140</sup> and retain nutrients and water that might otherwise be lost from the root zone.<sup>141,142</sup> Biochar is valuable when used as a fuel, carbon sink, or soil amendment<sup>143,145</sup> and a very promising technology for using it to make organic slow release nitrogenous fertilizer has been developed and patented.<sup>146</sup>

Biochar has increased soil pH and nutrient availability leading to crop yield improvements that persist for several years after a single application.<sup>147–152</sup> Nutrients contained in and applied with biochar materials can be responsible for short-term increases in crop growth.<sup>148</sup> The long-term improvement in soil fertility arises from the fact that the biomass thermal cracking process (pyrolysis) generates stable compounds consisting of single and condensed ring aromatic carbon with a high surface area per unit mass.<sup>153</sup> This surface becomes oxidized and cation exchange capacity (CEC) develops over time and can lead to greater nutrient retention in ‘aged’ as opposed to ‘fresh’ biochar.<sup>154–157</sup> The resulting high CEC presumably captures positively charged plant nutrients such as  $\text{NH}_4^+$ ,  $\text{K}^+$ ,  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  which are retained on the biochar surface and not lost through volatilization ( $\text{NH}_4^+ \rightarrow \text{NH}_3$ ) or leaching ( $\text{K}^+$ ,  $\text{Ca}^{2+}$  and possibly  $\text{Mg}^{2+}$ ).<sup>158</sup> The binding of  $\text{NH}_4^+$  to the biochar surface is of particular interest because this can slow the rate of nitrification ( $\text{NH}_4^+ \rightarrow \text{NO}_3^-$ ) and hence the loss of  $\text{N}_2\text{O}$  and  $\text{N}_2$  via denitrification. Biochar also binds  $\text{PO}_4^{3-}$  by surface adsorption,<sup>159</sup> thus providing a mechanism for better management of this key plant nutrient. Thus, biochar-amended soils may require less fertilizer to achieve target crop yields, leading to, for instance, less contamination of surface and ground water by  $\text{PO}_4^{3-}$  and  $\text{NO}_3^-$ <sup>160,161</sup> and less production of the greenhouse gas  $\text{N}_2\text{O}$ .<sup>162,163</sup>

Good quality biochar is very porous, contains less inorganic matter and can hold several times its weight in water.<sup>141,142</sup> Thus a field with  $10 \text{ t ha}^{-1}$  of added biochar might retain an additional  $30 \text{ t ha}^{-1}$  of water following a rainfall or irrigation event. This could be extremely important to crops growing in water-limited areas and

could also enhance the retention of N, S and P in soil and reduce fertilizer requirements.

The mean residence time (MRT) of microbially processed soil organic carbon is as short as 30 years.<sup>164</sup> In contrast, biochar creates a carbon pool with high stability.<sup>158,165</sup> MRTs of 6850 and 4035 years have been reported for biochar in the Amazonia Dark Earth region<sup>166,167</sup> and 364 years for purposefully applied biochar in the field, normalized for a mean annual temperature of  $10^\circ\text{C}$ .<sup>152</sup> Thus, biochar is very effective in the long-term sequestration of carbon into soils.

## Municipal solid waste as a source of bioethanol production

Municipal solid waste (MSW) consists of combustible and non-combustible wastes that come from household, municipal, commercial and industrial sites.<sup>168</sup> In the UK, 500 kg of waste per capita are produced each year amounting to 30 million tonnes of MSW annually.<sup>169</sup> These produce large amounts of GHGs at dumping sites and represent wasted energy.<sup>168</sup> Many countries including the US have increased their efforts to use MSW via recycling, thermo-chemical and biological conversions.

## Logistic challenges in using lignocellulosic feedstocks

The biomass-to-energy industry has been developing over the past 30 years<sup>194</sup> and during this period many challenges had to be overcome.<sup>195</sup> As the industry grows and matures, many other challenges arise, including logistics. Logistics is defined as ‘the art and science of obtaining, producing, and distributing material and product in the proper place and in proper quantities’<sup>198</sup> and plays a vital role in achieving operational excellence in all the types of industries, including bioenergy. To fully understand the importance of logistics in biomass, first it is required to understand the different processes that are involved (Fig. 1).

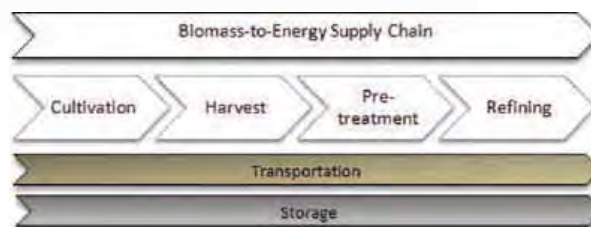


Figure 1. Biomass-to-energy supply chain

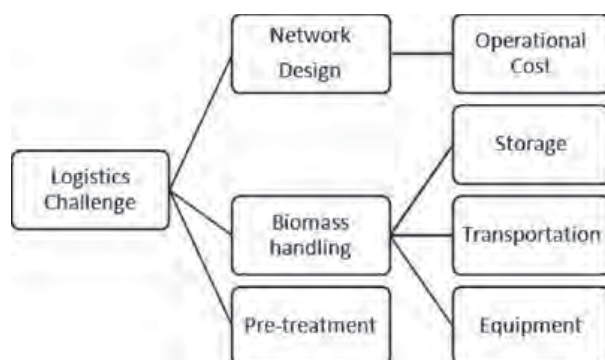


Figure 2. Logistic challenges

The biomass supply chain begins at cultivation and includes land use and crop selection which can significantly impact the expected yield and overall energy efficiency.<sup>197</sup> Additionally, site selection will impact the operational cost since the proximity of the feedstock to processing plants has a direct impact on transportation cost. The second step of the biomass-to-energy supply chain is harvest; it must consider the direct relationships between cultivation, storage, transportation and even conversion which are all significant drivers in logistics cost.<sup>9</sup> Harvesting will be discussed in the next section. Following harvest is pre-treatment where the first steps of biomass conversion occurs and where the feedstock is transformed for further downstream processing. During the pre-treatment process, typically only including physical processes, the properties of the biomass are modified using drying, densification and fractionation. From a logistic point of view, pre-treatment plays a key role in the supply chain; it helps to improve the storage and transportation process by increasing bulk density as well as improves the efficiency at the refining stage.

As many studies have described,<sup>196–201</sup> logistics is a major area of focus for development of the biomass-to-energy value chain and is critical when considering the wide spatial distribution of potential biomass, variable moisture content, low bulk density, and short harvesting window. The biggest challenges that need to be addressed in the short term include pre-treatment processing,<sup>197–204</sup> biomass handling (transportation and storage),<sup>196,197,199</sup> and network design<sup>200–202</sup> (Fig. 2).

AQ13

#### AQ14 Pre-treatment

##### Densification

Densification is the process that uses compression or compaction on the biomass to remove inter- and intra-particle

voids.<sup>209</sup> From the perspective of logistics, densification overcomes the low bulk density of the feedstock, which is a major barrier for developing biomass as an energy source.<sup>204</sup> Increasing the density of the feedstock directly impacts storage and transportation that are mainly based on the volume of biomass to be handled.<sup>197</sup> Densification allows for feedstock uniformity,<sup>204</sup> improving the handling efficiency, and process throughput (i.e. optimizing upload time, storage handling efficiency, etc.).

#### Biomass handling

Biomass handling is divided into two elements transportation and storage.

##### Transportation

Transportation accounts for as much as 35–60% of the biomass logistics cost.<sup>202</sup> Due to this major cost, optimizing transportation requirements and even small improvements in competitiveness can result in significant reductions in overall operation costs. The type of feedstock and method of pre-treatment will define the biomass bulk density and the maximum handling capacity using different transportation modules (truck, rail, ship, etc.)<sup>205</sup> and the steps required within the system, before reaching its destination. For example, truck loading and unloading operation cycle time can play an important role in transportation efficiency.<sup>196</sup> Sokhansanj and Hess<sup>196</sup> described that the loading of a 36-bale truck may take 30–40 min. By increasing the bulk density of the biomass, the handling efficiencies can be improved but the cost to reformat the biomass must be included in the final consideration.<sup>196</sup> Other factors such as trip distance, truck-carrying capacity and fossil fuel consumption can have a negative impact on local transport costs and logistics operations.<sup>197</sup> Ultimately, transport costs increase when the biomass feedstock is dispersed over large areas requiring significant road transport, but pre-treatment can be used to reduce these costs.<sup>197</sup>

##### Storage

Biomass storage and quality is directly impacted by its moisture content which can in turn directly impact the energy efficiency. For long-term storage of most bioenergy feedstock, the moisture should be below 17.5% on a dry weight basis. However, safe moisture content will depend on the selected feedstock.<sup>206</sup> Biomass with a moisture content between 40% and 60% is difficult to manage during storage; wet crops (<60% moisture) are more susceptible to microbial degradation and losses due to liquid effluent

production during storage.<sup>206,207</sup> Losses during wet storage are higher than dry storage and drying biomass below 20% moisture is required to avoid large losses during storage. Storage infrastructure can be a stronger determinant of biomass loss than moisture content for hay collected below 40% moisture.<sup>207</sup> Additionally, proper slotting efficiency (optimizing biomass storage according to its shape) is important and circular biomass bales are not the best option during storage; square bales have a better storage performance.<sup>196</sup>

## Network design

With many feedstock harvest locations the logistic operations tend to be a complex system that require a well-designed network and a robust transportation system to supply the bio-refinery. Based on biomass seasonality and network design, the feedstock may need to be stored in satellite locations for a period of time before its transformation into an energy source. The transportation cost will increase as a consequence of increasing distance and increased steps required before reaching the refining step. A poor network design will have a direct impact on the bio-refinery operational cost and consequently its financial performance.<sup>199</sup>

Most optimization models use the integrated biomass supply and logistic (IBSAL) software as the primary optimization tool.<sup>201,202</sup> IBSAL is a time-dependent simulation developed by the US Department of Energy as a tool for analysing and optimizing complex biomass supply systems.<sup>8</sup> IBSAL is a powerful tool for evaluating the supply chain from field to bio-refinery.<sup>201</sup>

Another technological tool that has been used for optimizing the biomass supply chain are geographical information systems (GIS), which enable the mapping of actual road networks.<sup>199</sup> Figure 3 is an example of how GIS can be used to define an optimal location for a large-scale



Figure 3. Road network definition by GIS model

bio-refinery and potential satellite storage locations based on biomass availability and network optimization.<sup>199</sup>

## Energy

Many studies have analyzed the energy consumption throughout the supply chain. Table 1 summarizes the different results from previous works mostly in wood feedstock. Dijkman and Benders<sup>211</sup> calculated the energy consumption of different processes in order to generate electricity from wood (willow and poplar; energy inputs for electricity from wood were given over a period of 20 years). The pre-treatment (drying and chipping) process is the most energy consuming in the supply chain because it is assumed that the initial moisture content is 50% and includes the process to reduce this moisture content to an optimal level of 15%. Transportation and fertilization are

**Table 1. Energy consumption throughout the supply chain. Dijkman and Benders<sup>211</sup> (a) Benoit *et al.*<sup>212</sup> calculated the energy consumed to generate 1 GJ of heat. (b) Valente *et al.*<sup>213</sup> Energy balance of a woody biomass supply chain (transportation assumed at 30 km). (c) Valente *et al.*<sup>214</sup> Primary energy input for 18 251 m<sup>3</sup> of woody biomass harvested, the transportation was assumed for 64 km**

Reference	Year	Feedstock	Unit	Energy Consumption					
				Cultivation	Harvest	Pre-treatment	Refining	Transport	Total
Dijkman & Benders	2010	Wood	MJ/ha	3,674	3,543	11,400	–		18,617
Benoit <i>et al.</i> (a)	2013	Eucalyptus	MJ/GJ	33.81	15.93		0.29	42.67	50
Benoit <i>et al.</i> (a)	2013	Eucalyptus	MJ/GJ	10.49	23.52		0.29	42.67	34
Valente <i>et al.</i> (b)	2011	Wood	K Wh/m <sup>3</sup>	10.46	4.22	17.8		11.92	32
Valente <i>et al.</i> (c)	2011	wood	MJ/m <sup>3</sup>	3.68	40.81	59.73		74.46	104



the two main drivers in the energy consumption for wood chips combustion; the model assumed transportation of 80 km for the wood chips in the calculation.<sup>212</sup>

## Biomass harvesting for lignocellulosic energy crops

Cheap and efficient harvesting methods, a critical part of the supply chain for using biomass as an energy source, depend on the type of feedstock to be used as a fuel source (wood, grasses, residues, or other materials). The first feedstocks used for biofuels to penetrate the market, such as crop grains and wood pellets, relied on harvesting technologies taken directly or modified from existing agricultural and forestry harvesting methods. Agricultural technologies included a mower used to cut the forage, a baler to densify and produce a transportable package and trucking for transportation. Additional equipment was required depending on the need to reduce the moisture content of the final product, including conditioners for crimping the material, a hay rake for windrowing or turning the grass, and a tedder to spread the grass. Traditional wood harvesting methods include a feller-buncher to cut and gather several trees at once, a skidder/forwarder to move logs from the forest to loading area, a loader/picker and truck/transport.<sup>171</sup>

AQ15

Additional equipment and or personnel may be required for delimiting and bucking the tree (cutting the log to size).

### Traditional harvesting technologies

#### Forestry

The challenge with using traditional harvesting methods is that the labor, energy consumption, and equipment costs do not always make sense from an economic or energy balance perspective when the end product is biofuels.

Alternative harvesting methods and equipment are slowly being developed and are focused on four major feedstocks for future biofuel development: purpose-grown grasses (switchgrass, *Miscanthus*, etc.), short rotation plantation species (willow and hybrid poplar), crop residues (corn-sorghum stover/cereal straw), and forestry residues.<sup>177–179</sup>

AQ16

Three systems described below are new technologies that have reduced the cost and energy requirements of harvesting forest residues, willow plantations and corn stover.

### New harvesting technologies

#### Biobaler

The Biobaler (Anderson Group Co., Chesterville, Quebec) was developed in 2005 as a willow harvesting system based

on an agricultural baler. The original system was designed to harvest plantation willows using a reinforced large round baler.<sup>177</sup> A cutter head was attached to the front of the round baler, allowing simultaneous cutting and conditioning of the woody material before ejecting it into the baler chamber for compression and wrapping. Subsequent variations of the original cutter included horizontal saw blades, flail hammers, flail knives, and a flail shredder. The various cutters were designed to operate on rough terrain and allow harvesting in natural brush conditions, with the cutter selected depending on the wood material to be harvested. The Biobaler unit has a full width cutting size of 2.6 m, and a length of up to 5.5 m requiring at least a 180 HP tractor to operate. The harvested bales weigh 500 to 600 kg, are 1.4 m in diameter (1.2 m wide) and have a density of 220 kg/m<sup>3</sup>, with 50% moisture content. The Biobaler offers a versatile alternative to harvest wild brush and planted woody crops. The technology is helpful for land management and provides a method to harvest otherwise neglected biomass. The biobaler, has an ability to harvest woody material up to 10 cm in diameter, 7 m in height and produces between 8 and 20 tonnes per hour (15 to 40 bales). It has operated in a range of harvesting locations including plantations, abandoned and fallow land, field edges, along roads, near watercourses, and understory harvesting. Management of the trees can be controlled through the selection of the cutters for the front end of the biobaler, with some offering a clean cut (saw blades) and others producing rough edges (flail cutting). Rough edges on stumps can allow water to stagnate and slowly penetrate the root system, resulting in detrimental stump and root health. Such slow regrowth of the woody crop after harvest is an advantage in some undergrowth management systems (abandoned or fallow land, river bank brush and understory vegetation in forests).

#### Willow harvester

Many willow harvesting systems are based on modifications to existing forage or sugar cane harvesters.<sup>179</sup> In Europe, the most effective machines are the modified Claas Jaguar corn harvester (Harsewinkel, Germany) and the Bender Harvester (Uppsala, Sweden). In North America, a Case-New Holland forage harvester (FX-45; Burr Ridge, IL, USA) has been modified for willow harvesting. The Claas harvester is a front head implement with saw blades, two blades per row that cuts stems 5 to 10 cm from the ground. The Bender willow harvester uses a single long, chain-saw cutting chain to cut the willow stems and cuts and chips the willow biomass in one pass using a 140HP tractor and harvester. The modified North

American system uses a modified row-independent CNH forage harvester to successfully harvest and chip willow biomass.<sup>179</sup> The unit has been tested on both plantation willow and hybrid poplar. Harvesting rates have been reported of up to 2 ha h<sup>-1</sup> with stems up to 13 cm diameter.

These modified forage harvesters cut the plants into chips, which are then blown into a wagon towed by the harvester or into a wagon/truck that drives alongside the harvester. A stem bundling system has been developed in Europe but has limited application for bioenergy production due to increased costs of transport and handling. However, bundled stems have improved results when used for transplanting of willow or when the biomass requires longer storage times.<sup>179</sup> Overall the forage harvesting equipment has been improving with harvest rates from 22 to 45 tons per hour (0.5 to 2 ha h<sup>-1</sup>).

### Stover/combine harvester

Corn/sorghum and cereal grain stover harvesting can occur in at least three forms: a single pass with simultaneous grain and stover harvesting, a two-pass grain harvesting and baling/forage harvesting of the stover windrow, and a three-pass system with grain harvesting, mowing/raking and baling/forage harvesting.<sup>178,182–188</sup> All three options have been used with varying levels of acceptance, and all use the combine for grain harvest. The single pass system consists of a modification to the combine with a corn stem cutting or ear-snapping header to provide a second stream of stover materials in addition to the grain.<sup>189</sup> This material is processed by a forage harvester type system that blows the material into a wagon for transport off the field. A modification to the single pass system is the two-pass system where the combine harvests the grain and provides a windrow of the stover for baling or forage harvesting in a second pass. The two-pass system is typically used for cereal grain with the baling of wheat and barley straw but is also used for corn stover.<sup>178</sup> The three-pass system uses the standard combine for corn seed harvest, then a mower cuts the stalks and windrows the material for baling or forage harvesting in the third step. Increased energy advantages, of up to 55%, have been reported with the single pass system.<sup>190</sup> Biomass removal efficiency has ranged from 35 to 93% depending on the type of corn head and head height used.<sup>189</sup> However, increased costs of the machinery and increased moisture content can limit its use. Placement of the stover into a windrow allows reduction in moisture content of the stover before baling, but can increase soil contamination of the stover.

The incorporation of agricultural and forestry harvesting methods into new biomass harvesting systems has

improved harvest efficiency, which can greatly reduce the time, cost, energy and manpower required for harvest. Reducing the machinery requirements and number of passes within the field/forest result in reduced labour costs and lower harvesting costs, allowing a lower cost biomass feedstock for downstream processes.

## Summary

The development of an energy source that is sustainable over the long term and will reduce our dependence on fossil fuels is beneficial both to the environment and the economy. Lignocellulosic feedstocks grown on land that is marginal for agriculture, using carefully selected species and production systems, have the potential to provide biofuels that are energy efficient, cost effective and environmentally sound. Ongoing research and development on suitable biomass feedstock is multifaceted and at scales that range from the cellular to the plantation. Genetic engineering is improving access to the carbohydrates stored in plant cell walls and reducing the energy costs currently associated with breaking down lignin. Perennial feedstocks that have a high productivity, strong persistence and wide adaptation to a variety of climatic and soil conditions are being selected and developed. Proper utilization of municipal solid waste can lead to economical biofuel production. Production systems that produce a maximized amount of fuel per unit of biomass, while maintaining crop sustainability with minimal inputs, are being developed. Enhanced nitrogen fixation and phosphorus uptake by plants and increased soil water and nutrient holding capacities are being developed and have the potential to improve soil fertility and allow lignocellulosic production systems to produce biofuels sustainably over the long term. Logistics play a vital role in achieving operational excellence in bioenergy and is a major area of focus for development of the biomass-to-energy value chain. Logistics are critical when considering the wide spatial distribution of potential biomass, variable moisture content, low bulk density, and short harvesting window. The biggest challenges that need to be addressed in the short term include pre-treatment processing, harvesting, transportation and network design. By way of recommendations we feel that:

1. We should develop plant genotypes with lignocellulosic materials that are effectively and efficiently converted into fluid fuels.
2. There is also a need to identify and genetically improve biofuel feedstock crops that are best adapted to specific geographical areas.



- There is a need to develop sustainable/low-input production methods that serve to enhance the yields of biofuel feedstock crops while also enhancing the energy balance and carbon life-cycle analysis of these crops.
- Effective, low energy and safe harvest and storage practices need to be developed for biofuel feedstock crops.

### Acknowledgments

This overview effort is the result of a Canada-California Strategic Initiative Project and received additional support from the McGill Network for Innovation in Biofuels and Bioproducts and BioFuelNet Canada. We thank Helen Fyles who assisted in compilation of the various materials.

### References

- Lovins AB and Rocky Mountain Institute, *Reinvesting Fire – Building Business Solutions for the New Energy Era*. Chelsea Green Publishing, White River Junction, Vermont (2011).
- Mabee WE and Saddler JN, Bioethanol from lignocellulose: Status and perspectives in Canada. *Bioresour Technol* **101**:4806–4813 (2010).
- Sanderson K, Lignocellulose: A chewy problem. *Nature* **474**:S12–S14 (2011).
- Kaufman RK and Shiers LD, Alternatives to conventional crude oil: When, how quickly, and market driven? *Ecol Econ* **67**:405–411 (2008).
- American Coalition for Ethanol, *Clean Air Facts*. [Online]. (2004). Available at: [www.ethanol.org](http://www.ethanol.org) [February 18, 2011].
- Searchinger T, Heimlich R and Houghton RA, Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. *Science* **319**:1238–1240 (2008).
- Pimentel D and Marklein A, Food versus biofuels: Environmental and economic costs. *Human Ecology* **37**:1–12 (2009).
- National Academy of Sciences (NAS), *Frontiers in Agricultural Research: Food, Health, Environment, and Communities*. [Online]. Washington, DC, NAS. (2004). Available at: [http://dels.nas.edu/rpt\\_briefs/frontiers\\_in\\_ag\\_final%20for%20print.pdf](http://dels.nas.edu/rpt_briefs/frontiers_in_ag_final%20for%20print.pdf) AQ18
- United States Department of Agriculture (USDA), *Major Land Uses*. [Online]. Economic Research Services, United States Department of Agriculture. (2007). Available at: <http://www.ers.usda.gov/data/majorlanduses> [March 12, 2011]. AQ19
- McLaughlin SB and Walsh ME, Evaluating environmental consequences of producing herbaceous crops for bioenergy. *Biomass Bioenergy* **14**:4317–324 (1998).
- Patzek TW, Thermodynamics of the corn–ethanol biofuel cycle. *Crit Rev Plant Sci* **23**:6519–6567 (2004).
- Srinivasan S, The food v. fuel debate: A nuanced view of incentive structures. *Renew Energy* **3**:950–954 (2009).
- Antizar-Ladislao B and Turrion-Gomez JL, Second-generation biofuels and local bioenergy systems. *Biofuels Bioprod Bioref* **2**:455–469 (2008).
- Stephanopoulos G, Metabolic engineering: Enabling technology for biofuels production. *Metab Eng* **10**:293–294 (2008).
- Hoogwijk M, Faaij A and Eickhout B, Potential of biomass energy out to 2100, for four IPCCsRES land-use scenarios. *Biomass Bioenergy* **29**:225–257 (2005).
- Dhugga KS, Maize biomass yield and composition for biofuels. *Crop Sci* **47**:2211–2227 (2007).
- Harris D and DeBolt S, Synthesis, regulation and utilization of lignocellulosic biomass. *Plant Biotech J* **8**:244–262 (2010).
- Gressel J, Transgenics are imperative for biofuel crops. *Plant Sci* **174**:246–264 (2008).
- Sticklen MB, Plant genetic engineering for biofuel production: Towards affordable cellulosic ethanol. *Nat Rev Genet* **9**:433–443 (2008).
- Reddy MS, Chen F, Shadle G, Jackson L, Aljoe H and Dixon RA, Targeted down-regulation of cytochrome P450 enzymes for forage quality improvement in alfalfa (*Medicago sativa* L.). *P Natl Acad Sci USA* **102**:16573–16578 (2005).
- Li L, Zhou YH, Cheng XF, Sun JY, Marita JM, Ralph J and Chiang VL, Combinatorial modification of multiple lignin traits in trees through multigene cotransformation. *P Natl Acad Sci USA* **100**:4939–4944 (2003).
- Weng J-K, Li X, Stout J and Chapple C, Independent origins of syringyl lignin in vascular plants. *P Natl Acad Sci USA* **105**:7887–7892 (2008).
- Chen F and Dixon RA, Lignin modification improves fermentable sugar yields for biofuel production. *Nat Biotechnol* **25**:759–761 (2007).
- Lin Y and Tanaka S, Ethanol fermentation from biomass resources: Current state and prospects. *Appl Microbiol Biot* **69**:627–642 (2006).
- Ziegler MT, Thomas SR and Danna DJ, Accumulation of a thermostable endo-1,4-β-D-glucanase in the apoplast of *Arabidopsis thaliana* leaves. *Mol Breeding* **6**:37–46 (2000).
- Biswas G, Ransom C and Sticklen M, Expression of biologically active *Acidothormus cellulolyticus* endoglucanase in transgenic maize plants. *Plant Sci* **171**:617–623 (2006).
- Oraby H, Venkatesh B, Dale B, Ahmad R, Ransom C, Oehmke J et al., Enhanced conversion of plant biomass into glucose using transgenic rice-produced endoglucanase for cellulosic ethanol. *Transgenic Res* **16**:739–749 (2007).
- Buanafina MM, Langdon T, Hauck B, Dalton S and Morris P, Expression of a fungal ferulic acid esterase increases cell wall digestibility of tall fescue (*Festuca arundinacea*). *Plant Biotech J* **6**:264–280 (2008).
- Sørensen HR, Meyer AS and Pedersen S, Enzymatic hydrolysis of water-soluble wheat arabinoxylan. 1. Synergy between α-l-arabinofuranosidases, endo-1,4-β-xylanases, and β-xylosidase activities. *Biotechnol Bioeng* **81**:726–731 (2003).
- Chatterjee A, Das NC, Raha S, Babbit R, Huang Q, Zaitlin D et al., Production of xylanase in transgenic tobacco for industrial use in bioenergy and biofuel applications. *In Vitro Cell Dev Biol Plant* **46**:198–209 (2010).
- Moon HS, Abercrombie JM, Kausch AP and Stewart CN, Sustainable use of biotechnology for bioenergy feedstocks. *Environ Manage* **46**:531–538 (2010).
- Wang Z-Y and Ge Y, Recent advances in genetic transformation of forage and turf grasses. *In Vitro Cell Dev Biol Plant* **42**:1–18 (2006).
- Sungrant Initiative, Feedstock Partnership. [Online]. Available at: <http://www.sungrant.org> [August, 2012]. AQ20
- Heaton E, Voigt T and Long SP, A quantitative review comparing the yields of two candidate C-4 perennial biomass crops in

- relation to nitrogen, temperature and water. *Biomass Bioenergy* **27**:21–30 (2004).
35. Parrish DJ and Fike JH, The biology and agronomy of switchgrass for biofuels. *Crit Rev Plant Sci* **24**:423–459 (2005).
36. Perrin RK, Vogel KP, Schmer MR and Mitchell RB, Farm-scale production cost of switchgrass for biomass. *Bioenergy Res* **1**:91–97 (2008).
37. Sanderson MA, Adler PR, Boateng AA, Casler MD and Sarath G, Switchgrass as a biofuels feedstock in the USA. *Can J Plant Sci* **86**:1315–1325 (2006).
38. Jacob K, Zhou F and Paterson AH, Genetic improvement of C<sub>4</sub> grasses as cellulosic biofuel feedstocks. *In Vitro Cell Dev Biol Plant* **45**:291–305 (2009).
39. Heaton EA, Boersma NN, Caveny JD, Voigt TB and Dohleman FG, Miscanthus for biofuel production. eXtension Bioenergy Feedstock Community of Practice (2010).
- AQ21 40. Somerville C, Youngs H, Taylor C, Davis SC and Long SP, Feedstocks for lignocellulosic biofuels. *Science* **329**:790–792 (2010).
41. Carroll A and Somerville C, Cellulosic biofuels. *Annu Rev Plan Biol* **60**:165–182 (2009).
42. Wrobel C, Coulman BE and Smith DL, The potential use of reed canarygrass (*Phalaris arundinacea* L.) as a biofuel crop. *Acta Agr Scand B-S P* **59**:1–18 (2008).
43. Barney JN and Ditomaso JM, Bioclimatic predictions of habitat suitability for the biofuel switchgrass in North America under current and future climate scenarios. *Biomass Bioenergy* **34**:124–133 (2010).
44. Wright L and Turhollow A, Switchgrass selection as a “model” bioenergy crop: A history of the process. *Biomass Bioenergy* **34**:851–868 (2010).
45. Hultquist SJ, Vogel KP, Lee KJ, Arumuganathan K and Kaeppeler S, DNA content and chloroplast DNA polymorphisms among switchgrasses from remnant midwestern prairies. *Crop Sci* **37**:595–598 (1997).
46. Missaoui AM, Paterson AH and Bouton JH, Investigation of genomic organization in switchgrass (*Panicum virgatum* L.) using DNA markers. *Theor Appl Genet* **110**:1372–1383 (2005).
47. Casler MD, Stendal CA, Kapich L and Vogel KP, Genetic diversity, plant adaptation regions, and gene pools for switchgrass. *Crop Sci* **47**:2261–2273 (2007).
48. Bhandari H, Saha M, Fasoula V and Bouton J, Heritability of biomass yield and yield components in lowland switchgrass. *Proceedings Annual Meeting of the American Society of Agronomy*. Long Beach, CA (2010).
- AQ22 49. Boe A and Lee DK, Genetic variation for biomass production in prairie cordgrass and switchgrass. *Crop Sci* **47**:929–934 (2007).
50. Boe A, Variation between two switchgrass cultivars for components of vegetative and seed biomass. *Crop Sci* **47**:634–640 (2007).
51. Boe A and Beck DL, Yield components of biomass in switchgrass. *Crop Sci* **48**:1306–1311 (2008).
52. Martinez-Reyna JM and Vogel KP, Heterosis in switchgrass: Spaced plants. *Crop Sci* **48**:1312–1320 (2008).
53. Vogel KP and Mitchell RB, Heterosis in switchgrass: Biomass yield in swards. *Crop Sci* **48**:2159–2164 (2008).
54. Bouton J, Molecular breeding of switchgrass for use as a biofuel crop. *Curr Opin Genet Dev* **17**:553–558 (2007).
55. Vogel K, Switchgrass, in *Warm-Season (C4) Grasses*, ed by Moser L, Burson B and Sollenberger L. American Society of Agronomy/Crop Science Society of America/Soil Science Society of America, Madison, Wisconsin, pp. 561–588 (2004).
56. Cassida KA, Muir JP, Hussey MA, Read JC, Venuto BC and Ocumpaugh WR, Biomass yield and stand characteristics of switchgrass in south central US environments. *Crop Sci* **45**:673–681 (2005).
57. Gustafson DM, Boe A and Jin Y, Genetic variation for Puccinia emaculata infection in switchgrass. *Crop Sci* **43**:755–759 (2003).
58. Hopkins AA, Vogel KP, Moore KJ, Johnson KD and Carlson IT, Genotypic variability and genotype X environment interactions among switchgrass accessions from the midwestern USA. *Crop Sci* **35**:565–571 (1995).
59. Boydston RA, Collins HP and Fransen SC, Response of three switchgrass (*Panicum virgatum*) cultivars to mesotrione, quinclorac, and pendimethalin. *Weed Technol* **24**:336–341 (2010).
60. Fike JH, Parrish DJ, Wolf DD, Balasko JA, Green JT, Rasnake M and Reynolds JH, Switchgrass production for the upper southeastern USA: Influence of cultivar and cutting frequency on biomass yields. *Biomass Bioenergy* **30**:207–213 (2006).
61. El-Nashaar HM, Banowetz GM, Griffith SM, Casler MD and Vogel KP, Genotypic variability in mineral composition of switchgrass. *Bioresource Technol* **100**:1809–1814 (2009).
62. Yan JH, Hu ZJ, Pu YQ, Brummer EC and Ragauskas AJ, Chemical compositions of four switchgrass populations. *Biomass Bioenergy* **34**:48–53 (2010).
63. Heaton EA, Dohleman FG and Long SP, Seasonal nitrogen dynamics of *Miscanthus x giganteus* and *Panicum virgatum*. *Glob Change Biol Bioenergy* **1**:297–307 (2009).
64. Hodkinson TR, Chase MW and Renvoize SA, Genetic resources of *Miscanthus*. *Aspect Appl Biol* **65**:239–248 (2001).
65. Rayburn AL, Crawford J, Rayburn CM and Juvik JA, Genome size of three *Miscanthus* species. *Plant Mol Bio Rep* **27**:184–188 (2008).
66. Hodgson EM, Lister SJ, Bridgwater AV, Clifton-Brown J and Donnison IS, Genotypic and environmentally derived variation in the cell wall composition of *Miscanthus* in relation to its use as a biomass feedstock. *Biomass Bioenergy* **34**:652–660 (2010).
67. Barney JN and Ditomaso JM. Nonnative species and bioenergy: Are we cultivating the next invader? *Bioscience* **58**:64–70 (2008).
68. Quinn L, Allen D and Stewart JR, Invasiveness potential of *Miscanthus sinensis*: Implications for bioenergy production in the United States. *GCB Bioenergy* **2**:310–320 (2010).
69. Clifton-Brown JC and Lewandowski I, Overwintering problems of newly established *Miscanthus* plantations can be overcome by identifying genotypes with improved rhizome cold tolerance. *New Phytol* **148**:287–294 (2000).
70. Jorgenson U, Mortensen J, Kjedsen JB and Schwarz KU, Establishment, development and yield quality of fifteen *Miscanthus* genotypes over three years in Denmark. *Acta Agr Scand B-S P* **53**:190–199 (2003).
71. Atienza SG, Satovic Z, Petersen KK, Dolstra O and Martin A, Identification of QTLs influencing agronomic traits in *Miscanthus sinensis* Anders. I. Total height, flag-leaf height and stem diameter. *Theor Appl Genet* **107**:123–129 (2003).
72. de Cesare M and Hodkinson TR, Chloroplast DNA markers (cpSSRs, SNPs) for miscanthus, saccharum and related grasses (panicoideae, poaceae). *Mol Breeding* **26**:539–544 (2010).



73. Heaton EA, Dohleman FG and Long SP, Meeting US bio-fuel goals with less land: The potential of *Miscanthus*. *Glob Change Biol* **14**:2000–2014 (2008).
74. Heaton EA, Dohleman FG, Miguez AF Juvik JA, Lozovaya V, Widholm J *et al.*, *Miscanthus*: A promising biomass crop. *Adv Bot Res* **56**:75–137 (2010).
75. Jones MB and Walsh M, *Miscanthus for Energy and Fibre*. James and James Ltd, London (2001).
76. Long SP, Environmental responses, in: *C4 Plant Biology*, ed by Sage R and Monson RK. Kluwer Academic Press (1999).
77. Miguez FE, Zhu XG, Humphries S, Bollero GA and Long SP, A semimechanistic model predicting the growth and production of the bioenergy crop *Miscanthus x giganteus*: description, parameterization and validation. *Glob Change Biol Bioenergy* **1**:282–296 (2009).
78. Pyter R, Heaton EA, Dohleman FG, Voigt TB and Long SP, Agronomic experiences with *Miscanthus x giganteus* in Illinois, in *Biofuels Methods and Protocol*, ed by Mielenz JR. The Humana Press, New York, NY (2009).
79. Jain AK, Khanna M, Erickson M and Huang HX, An integrated biogeochemical and economic analysis of bioenergy crops in the Midwestern United States. *Glob Change Biol Bioenergy* **2**:217–234 (2010).
80. Miguez FE, Villamil MB, Long SP and Bollero GA, Meta-analysis of the effects of management factors on *Miscanthus x giganteus* growth and biomass production. *Agr Forest Meteorol* **148**:1280–1292 (2008).
81. Kim S, Lane RA and Lee DK, Genome size and chromosome analyses in prairie cordgrass. *Crop Sci* **50**:2277–2282 (2010).
82. Boe A, Owens V, Gonzalez-Herandez J, Stein J, Lee DK and Koo BC, Morphology and biomass production of prairie cordgrass on marginal lands. *GCB Bioenergy* **1**:240–250 (2009).
83. Stanton BJ, Neale DB and Shanwen L, *Populus* breeding: From the classical to the genomic approach. *Plant Genet Genom Crop Model* **8**:309–348 (2010).
84. Boerjan W, Biotechnology and the domestication of forest trees. *Curr Opin Biotech* **16**:159–166 (2005).
85. Frewen BE, Chen TH, Howe GT, Davis J, Rohde A, Boerjan W and Bradshaw HD, Quantitative trait loci and candidate gene mapping of bud set and bud flush in *Populus*. *Genetics* **154**:837–845 (2000).
86. Wegrzyn JL, Eckert AJ, Choi M, Lee JM, Stanton BJ, Sykes R *et al.*, Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (*Populus trichocarpa*, *Salicaceae*) Secondary Xylem. *New Phytol* **188**:515–532 (2010).
87. Fillatti JJ, Sellmer J, Mccown B, Haissig B and Comai L, Agrobacterium mediated transformation and regeneration of populus. *Mol Gen Genet* **206**:192–199 (1987).
88. Rowe RL, Street NR and Taylor G, Identifying potential environmental impacts of large-scale deployment of dedicated bioenergy crops in the UK. *Renew Sust Energy Rev* **13**:271–290 (2007).
89. Kopp RF, White EH, Abrahamson LP, Nowak CA, Zsuffa L and Burn KF, Willow biomass trials in Central New York State. *Biomass Bioenergy* **5**:179–187 (1993).
90. Brown KR and Van Den Driessche R, Effects of nitrogen and phosphorus fertilization on the growth and nutrition of hybrid poplars on Vancouver Island. *New Forest* **29**:89–104 (2005).
91. Peichl M, Thevathasan NV, Gordon AM, Huss J and Abohassan RA, Carbon sequestration potentials in temperate tree-based intercropping systems in southern Ontario, Canada. *Agroforest Syst* **66**:243–257 (2006).
92. Montagnini F and Nair PK, Carbon sequestration: An under-exploited environmental benefit of agroforestry systems. *Agroforest Syst* **61**:281–295 (2004).
93. Bambrick AD, Whalen JK, Bradley RL, Cogliastro A, Gordon AM, Olivier A and Thevathasan NV, Spatial heterogeneity of soil organic carbon in tree-based intercropping systems in Quebec and Ontario, Canada. *Agroforest Syst* **79**:343–353 (2010).
94. Labrecque M and Teodorescu TI, Field performance and biomass production of 12 willow and poplar clones in short-rotation coppice in southern Quebec (Canada). *Biomass Bioenergy* **29**:1–9 (2005).
95. Welham C, Van Rees K, Seely B and Kimmins H, Projected long-term productivity in Saskatchewan hybrid poplar plantations: Weed competition and fertilizer effects. *Can J Forest Res* **37**:356–370 (2007).
96. Zhang SY, Yu Q, Chauret G and Koubaa A, Selection for both growth and wood properties in hAKrid poplar clones. *Forest Sci* **49**:901–908 (2003).
97. Luo X-B and Polle A, Wood composition and energy content in a poplar short rotation plantation on fertilized agricultural land in a future CO<sub>2</sub> atmosphere. *Glob Change Biol* **15**:38–47 (2009).
98. Dahlberg J, Berenji J, Sikora V and Latkovic D, Assessing sorghum [*Sorghum bicolor* (L) Moench] germplasm for new traits: Food, fuels & unique uses. *Maydica* **56**(1750):85–92 (2011).
99. Patterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H *et al.*, The sorghum bicolor genome and the diversification of grasses. *Nature* **457**:151–156 (2009).
100. Menossi M, Silva-Filho MC, Vincentz M, Van-Sluys MA, Souzam GM, Sugarcane functional genomics: Gene Discovery for agronomic trait development. *Int J Plant Genomics* **458732** (2008).
101. Wang J, Roe B, Macmil S, Yu Q, Murray JE, Tang H *et al.*, Microcollinearity between autopolyploid sugarcane and diploid sorghum genomes. *BMC Genomics* **11**: 261 (2010).
102. Marsalis MA, Angadi SV, Contreras-Govea FE. Dry matter yield and nutritive value of corn, forage sorghum, and BMR forage sorghum at different plant populations and nitrogen rates. *Field Crop Res* **116**:52–57 (2010).
103. Rocateli AC, *A new spin on an old crop for bioenergy: Sorghum*. An MSc thesis, Auburn University (2010).
104. Gray EJ and Smith DL, Intracellular and extracellular PGPR: Commonalities and distinctions in the plant-bacterium signaling processes. *Soil Biol Biochem* **37**:395–412 (2005).
105. Whipps JM, Microbial interactions and biocontrol in the rhizosphere. *J Exp Bot* **52**:487–511 (2001).
106. Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA *et al.*, The path forward for biofuels and biomaterials. *Science* **311**:484–489 (2006).
107. Souleimanov A, Prithviraj B and Smith DL, The major nod factor of *Bradyrhizobium japonicum* promotes early growth of soybean and corn. *J Exp Bot* **53**:1929–1934 (2002).
108. Prithviraj B, Zhou X, Souleimanov A, Khan WK and Smith DL, A host specific bacteria-to-plant signal molecule (nod factor) enhances germination and early growth of diverse crop plants. *Planta* **216**:437–445 (2003).

AQ23

AQ24

109. Almaraz J, Zhou X and Smith DL, Gas exchange characteristics and dry matter accumulation of soybean treated with Nod factors. *J Plant Physiol* **164**:1391–1393 (2007).
110. Dyachok JV, Tobin AE, Price NPJ and von Arnold S, Rhizobial nod factors stimulate somatic embryo development in *Picea abies*. *Plant Cell Rep* **19**:290–297 (2000).
111. Dyachok JV, Wiweger M, Kenne L and von Arnold S, Endogenous nod-factor-like signal molecules promote early somatic embryo development in Norway spruce. *Plant Physiol* **128**:523–533 (2002).
112. Olah B, Briere C, Becard G, Denarie J and Gough C, Nod factors and a diffusible factor from arbuscular mycorrhizal fungi stimulate lateral root formation in *Medicago truncatula* via the DMI1 / DMI2 signalling pathway. *Plant J* **44**:195–207 (2005).
113. Chen C, McIver J, Yang Y, Bai Y, Schultz B and McIver A, Foliar application of lipo-chitooligosaccharides (nod factors) to tomato (*Lycopersicon esculentum*) enhances flowering and fruit production. *Can J Plant Sci* **87**:365–72 (2007).
114. Duzan HM, Mabood F, Souleimanov A and Smith DL, Nod factor induces soybean resistance to powdery mildew. *Plant Physiol Bioch* **43**:1022–1030 (2005).
115. Savouré A, Magyar Z, Pierre M, Brown S, Schultze M, Dudits D et al., Activation of the cell cycle machinery and the isoflavonoid biosynthesis pathway by active *Rhizobium meliloti* nod signal molecules in *Medicago microcallus* suspensions. *EMBO J* **13**:1093–1102 (1994).
116. Bai Y, Pan B, Charles T and Smith D, Coinoculation dose and root zone temperature for plant growth promoting rhizobacteria on soybean [*Glycine max* (L.) Merr]. *Soil Biol Biochem* **34**:1953–1957 (2002).
117. Bai Y, Zhou X and Smith DL, Enhanced soybean plant growth due to coinoculation of *Bacillus* strains with *Bradyrhizobium japonicum*. *Crop Sci* **43**:1774–1781 (2003).
118. Gray EJ, Lee KD, Souleimanov AM, Di Falco MR, Zhou X, Ly A et al., A novel bacteriocin, Thuricin 17, produced by PGPR strain *Bacillus thuringiensis* NEB17: isolation and classification. *J Appl Microbiol* **100**:545–554 (2006).
119. Gray E, Di Falco M, Souleimanov A and Smith DL, Proteomic analysis of the bacteriocin, thuricin 17 produced by *Bacillus thuringiensis* NEB17. *FEMS Microbiol Lett* **255**:27–32 (2006).
120. Lee KD, Gray EJ, Mabood F, Jung WJ, Charles T, Clark SR et al., The class IId bacteriocin thuricin 17 increases plant growth. *Planta* **229**:747–755 (2009).
121. Smith D, Lee KD, Gray E, Souleimanov A, Zhou X and Charles T, Thuricin 17 for promoting plant growth and disease resistance and transgenic plants. PCT/CA2008/000921 (2008).
122. Oscariz JC, Lasa I and Pisabarro AG, Detection and characterization of cerein 7, a new bacteriocin produced by *Bacillus cereus* with a broad spectrum of activity. *FEMS Microbiol Lett* **178**:337–341 (1999).
123. Kamoun F, Mejdoub H, Aouissaoui H, Reinbolt J, Hammami A and Jaoua S, Purification, amino acid sequence and characterization of Bac Thuricin F4, a new bacteriocin produced by *Bacillus thuringiensis*. *J Appl Microbiol* **98**:881–888 (2005).
124. Parret AHA and De Mot R, Bacteria killing their own kind: Novel bacteriocins of pseudomonas and other c-proteobacteria. *Trends Microbiol* **10**:107–112 (2002).
125. Oresnick IJ, Twelker S and Hynes MF, Cloning and characterization of a *Rhizobium leguminosarum* gene encoding a bacteriocin with similarities to RTX toxins. *Appl Environ Microb* **65**:2833–2840 (1999).
126. Wilson RA, Handley BA and Beringer JE, Bacteriocin production and resistance in a field population of *Rhizobium leguminosarum* biovar viciae. *Soil Biol Biochem* **30**:413–417 (1998).
127. Glick BR, The enhancement of plant growth by free-living bacteria. *Can J Microbiol* **41**:109–117 (1995).
128. Vessey JK, Plant growth promoting rhizobacteria as biofertilizers. *Plant Soil* **255**:571–586 (2003).
129. Döbereiner J, Biological nitrogen fixation in the tropics: Social and economic contributions. *Soil Biol Biochem* **29**:771–774 (1997).
130. Pessoa Jr A, Roberto IC, Menossi M, Santos RR, Ortega Filho S, Penna TC, Perspectives on bioenergy and biotechnology in Brazil. *Appl Biochem Biotechnol* **121**:59–70 (2005).
131. Boddey RM, Pereira JA, Hungria M, Thomas RJ and Neves MC, Methods for the study of nitrogen assimilation and transport in grain legumes. *MIRCEN J Appl Microb* **3**:3–22 (1987).
132. Dobbelaere S, Vanderleyden J and Okon Y, Plant growth-promoting effects of diazotrophs in the rhizosphere. *Crit Rev Plant Sci* **22**:107–149 (2003).
133. Cocking EC, Stone PJ and Davey MR, Intracellular colonization of roots of *Arabidopsis* and crop plants by *Gluconacetobacter diazotrophicus*. *In Vitro Cell Dev Biol Plant* **42**:74–82 (2006).
134. James EK, Gyaneshwar P, Mathan N, Barraquio WL, Reddy PM, Iannetta P et al., Infection and colonization of rice seedlings by the plant growth-promoting bacterium *Herbaspirillum seropedicae* Z67. *Mol Plant Microbe In* **15**:894–906 (2002).
135. Malik KA, Bilal R, Mehnaz S, Rasul G, Mirza MS and Ali S, Association of nitrogen-fixing, plant-growth-promoting rhizobacteria (PGPR) with kallar grass and rice. *Plant Soil* **194**:37–44 (1997).
136. Bonfante P and Genre A, Plants and arbuscular mycorrhizal fungi: An evolutionary-developmental perspective. *Trends Plant Sci* **13**:492–498 (2008).
137. Harrison MJ, Molecular and cellular aspects of the arbuscular mycorrhizal symbiosis. *Ann Rev Plant Biol* **50**:361–389 (1999).
138. Bucher M, Functional biology of plant phosphate uptake at root and mycorrhiza interfaces. *New Phytol* **173**:11–26 (2007).
139. Antunes PM, Koch AM, Dunfield KE, Hart MM, Downing A, Rillig MC et al., Influence of commercial inoculation with *Glomus intraradices* on the structure and functioning of an AM fungal community from an agricultural site. *Plant Soil* **317**:257–266 (2009).
140. Downie A, Klatt P, Downie R and Munroe P, Slow pyrolysis: Australian demonstration plant successful on multifeedstocks, in *Bioenergy 2007 Conference*. Finland (2007).
141. Novak JM, Busscher WJ, Laird DL, Ahmedna M, Watts DW and Niandou MAS, Impact of biochar amendment on fertility of a southeastern coastal plain soil. *Soil Sci* **174**:105–112 (2009).
142. Major J, Steiner C, Downie A and Lehmann J, *Biochar for Environmental Management: Science and Technology*. Earthscan, London, UK (2009).
143. Briens C, Piskorz J and Berruti F, Biomass valorization for fuel and chemicals production – A review. *Intl J Chem R Eng* **6**:R2– (2008).
144. Glaser B, Prehistorically modified soils of central Amazonia: A model for sustainable agriculture in the twenty-first century. *Philos T Roy Soc B* **362**:187– (2007).



145. Lehmann J, Gaunt J and Rondon M, Bio-char sequestration in terrestrial ecosystems - a review. *Mitig Adapt Strat Glob Change* **11**:403– (2006).
146. Radlein DAG, Piskorz J and Majerski P, Method of producing slow-release nitrogenous organic fertilizer from biomass. US Patent 5,676,727 (1997).
147. Glaser B, Lehmann J and Zech W, Ameliorating physical and chemical properties of highly weathered soils in the tropics with charcoal - a review. *Biol Fert Soils* **35**:219– (2002).
148. Lehmann J, da Silva JP, Steiner C, Nehls T, Zech W and Glaser B, Nutrient availability and leaching in an archaeological Anthrosol and a Ferralsol of the Central Amazon basin: Fertilizer, manure and charcoal amendments. *Plant Soil* **249**:343– (2003).
149. Lehmann J, da Silva Jr JP, Rondon M, Cravo MS, Greenwood J, Nehls T *et al.*, Slash-and-char - a feasible alternative for soil fertility management in the central Amazon? Paper No. 449. *17th World Congress of Soil Science*, Bangkok, Thailand (2002).
- AQ28 150. Rondon MA, Lehmann J, Ramirez J and Hurtado M, Biological nitrogen fixation by common beans (*Phaseolus vulgaris* L.) increases with bio-char additions. *Biol Fert Soils* **43**: 699– (2007).
151. Steiner C, Glaser B, Teixeira WG, Lehmann J, Blum WE and Zech W, Nitrogen retention and plant uptake on a highly weathered central Amazonian Ferralsol amended with compost and charcoal. *J Plant Nutr Soil SC* **171**:893– (2008).
152. Major J, Lehmann J, Rondon M and Goodale C, Fate of soil-applied black carbon: Downward migration, leaching and soil respiration. *Glob Change Biol* **16**:1366–1379 (2010).
153. Lehmann J, A handful of carbon. *Nature* **447**:143– (2007).
154. Liang B, Lehmann J, Solomon D, Kinyangi J, Grossman J, O'Neill B *et al.*, Black carbon increases cation exchange capacity in soils. *Soil Sci Soc Am J* **70**:1719– (2006).
155. Cheng C-H, Lehmann J and Engelhard MH, Natural oxidation of black carbon in soils: Changes in molecular form and surface charge along a climosequence. *Geoderma Cosmochim Acta* **72**:1598– (2008).
- AQ29 156. Cheng C-H, Lehmann J, Thies JE and Burton SD, Stability of black carbon in soils across a climatic gradient. *J Geophys Res-Biogeosci* **113**:GO2027 (2008).
157. Cheng C-H, Lehmann J, Thies JE, Burton SD and Engelhard MH, Oxidation of black carbon by biotic and abiotic processes. *Org Geochem* **37**:1477– (2006).
158. Glaser B, Haumaier L, Guggenberger G and Zech W, The 'Terra Preta' phenomenon: A model for sustainable agriculture in the humid tropics. *Naturwissenschaften* **88**:37– (2001).
159. Lehmann J, Lan Z, Hyland C, Sato S, Solomon D and Ketterings Q, Long-term dynamics of phosphorus forms and retention in manure-amended soils. *Environ Sci Technol* **39**:6672–6680 (2005).
160. Almasri MN and Kaluarachchi JJ, Assessment and management of long-term nitrate pollution of ground water in agriculture-dominated watersheds. *J Hydrol* **295**:225–245 (2004).
161. Elmi A, Madani A, Gordon R, MacDonald R and Stratton GW, Nitrate nitrogen in the soil profile and drainage water as influenced by manure and mineral fertilizer application in a barley-carrot production system. *Water Air Soil Poll* **160**:119–132 (2005).
162. Almaraz JJ, Mabood F, Zhou X, Strachan I, Ma B and Smith DL, Performance of agricultural systems under contrasting growing season conditions in south-western Quebec. *J Agron Crop Sci* **195**:319–327 (2009).
163. Van Zwielen L, Singh B, Joseph S, Kimber S, Cowie A and Chan Y, Biochar and emissions of non-CO<sub>2</sub> greenhouse gases from soil (Chapter 13), in *Biochar for Environmental Management: Science and Technology*, ed by Lehmann J and Joseph S. Earthscan, London, UK, pp. 227– (2009).
164. Marschner B, Brodowski S, Dreves A, Gleixner G, Gude A, Grootes PM *et al.*, How relevant is recalcitrance for the stabilization of organic matter in soils? *J Plant Nutr Soil Sci* **171**:91–110 (2008).
165. Forbes MS, Raison RJ and Skjemstad JO, Formation, transformation and transport of black carbon (charcoal) in terrestrial and aquatic ecosystems. *Sci Total Environ* **370**:190– (2006).
166. German LA, Historical contingencies in the coevolution of environment and livelihood: Contributions to the debate on Amazonian Black Earth. *Geoderma* **111**:307– (2003).
167. Liang B, Lehmann J, Solomon D, Sohi S, Thies JE, Skjemstad JO *et al.*, Stability of biomass-derived black carbon in soils. *Geochim Cosmochim Acta* **72**:6069– (2008).
168. Dalai AK, Batta N, Eswaremoorthi I and Schoenau GJ, Gasification of refuse derived fuel in a fixed bed reactor for syngas production. *Waste Manage* **29**:252–258 (2009).
169. Hernandez-Atonel FD, Ryu C, Sharifi, VN and Swithenbank J, Combustion of refuse-derived fuel in a fluidized bed. *Chem Eng Sci* **62**:627–635 (2007).
170. Naik SN, Goud VV, Rout PK and Dalai AK, Production of first and second generation biofuels: A comprehensive review. *Renew Sust Rev* **14**:578–597 (2010).
171. Drushka K and Kontinen H, *Tracks in the Forest: The Evolution of Logging Machinery*. Timberjack Group, Helsinki, distributed by Harbour Publishing, Madeira, BC (1997).
172. MacBain R, *Pelleting Animal Feed*. American Feed Manufacturers Assoc, Arlington, VA (1996).
173. Samson R, Mani S, Boddey R, Sokhansanj S, Quesada D, Urquiaga S *et al.*, The potential of C4 perennial grasses for developing a global BIOHEAT industry. *Crit Rev Plant Sci* **24**:461–494 (2005).
174. Mislevy P and Fluck RC, Harvesting, baling and energetics of tall grass for biomass energy production: A case study, in *Energy from Biomass and Wasters XVI*, ed by Klass DL. Institute of Gas Technology, Chicago, IL (1993).
175. Ho GE, Crop residues - II. How much can be economically harvested? *Biomass* **7**:199–214 (1985).
176. Rider AR and Barr SD, *Fundamentals of Machine Operation - Hay and Forage Harvesting*. John Deere Service Publications, Moline, Illinois (1976).
177. Lavoie F, Savoie P, D'Amours L and Joannis H, Development and field performance of a willow cutter-shredder-baler, *International Conference on Crop Harvesting and Processing*, 11–14 February 2007, ASABE Publication, Number 701P0307e, Louisville, KY, USA, (2007).
178. Shinnors KJ, Binversie BN and Savoie P, Whole-plant corn harvesting for biomass: Comparison of single-pass and multi-pass harvest systems. *American Society of Agricultural Engineers Paper 036089*. ASAE, St Joseph, MI (2003).
179. Volk TA, Abrahamson LP, Pripke E, Aneshansley D, White EH and Smart LB, Development of a willow biomass crop harvester based on a New Holland forage harvester and specially designed willow cutting head. *Society of American Foresters Agroforestry Working Group Newsletter* **5**, February 1, (2008).



180. Busse MD, Simon SA and Riegel GM, Tree-growth and understory responses to low-severity prescribed burning in thinned *Pinus ponderosa* forests of Central Oregon. *Forest Science* **46**:258–268 (2000).
181. Windell K and Bradshaw S, *Understory Biomass Reduction Methods and Equipment Catalog*, Tech Rep 0051-2826-MTDC. US Department of Agriculture, Forest Service, Missoula Technology Development Center, Missoula, MT, pp.156 (2000).
182. Albert WW and Stephens LE, Stalklage silage harvested with a converted combine, *American Society of Agricultural Engineers Paper 69313*. ASAE, St Joseph, MI (1969).
183. Ayres GE and Buchele WF, Harvesting and storing corn plant forage. *American Society of Agricultural Engineers Paper 71665*. ASAE, St Joseph, MI (1971).
184. Ayres GE and Buchele WF, An evaluation of machinery systems for harvesting corn plant forage. *American Society of Agricultural Engineers Paper 761015*. ASAE, St Joseph, MI (1976).
185. Burgin KH. Corn and stalk harvester. US Patent 2,385,193 (1941).
186. Buchele WF, Research in developing more efficient harvesting machinery and utilization of crop residues. *T Am Soc Agr Eng* **19**:809–811 (1976).
187. Hitzhusen TE, Marley SJ and Buchele WF, Beefmaker II: Developing a total corn harvester. *Agr Eng* **51**:632–634 (1970).
188. Shinnors KJ, Boettcher GC, Hoffman DS, Munk JT, Muck RE and Weimer PJ, Single-pass harvest of corn grain and stover: Performance of three harvester configurations. *Trans ASABE* **52**:51–60 (2009).
189. Shinnors KJ, Adsit GS, Binversie BN, Digman MF, Muck RE and Weimer PJ, Single-pass, split stream harvest of corn grain and stover. *Trans ASABE* **50**:355–363 (2007).
190. Shinnors KJ, Binversie BN, Muck RE and Weimer PJ, Comparison of wet and dry corn stover harvest and storage. *Biomass Bioenerg* **3**:211–221 (2007).
- AQ30 191. CenUSA Bioenergy Annual Report. [Online]. Iowa State University (2013). Available at: [https://www.cenusa.iastate.edu/PublicFile/\\_GetPublicFile?publicFileId=55](https://www.cenusa.iastate.edu/PublicFile/_GetPublicFile?publicFileId=55)
192. Wilson DM, Dalluge DL, Rover M, Heaton EA and Brown RC, Crop management impacts biofuel quality: Influence of switchgrass harvest time on yield, nitrogen, and ash of fast pyrolysis products. *Bioenerg Res* **6**:103–113 (2013).
193. James LK, Swinton SM and Thelen KD, Profitability analysis of cellulosic energy crops compared with corn. *Agron J* **102**:675–687 (2010).
194. Center for Energy, Biomass Timeline [Online]. Available at: [www.centreforenergy.com](http://www.centreforenergy.com) [April 3, 2013].
195. Electrical Power Research Institute, Bio power generation: Biomass issues, fuels, technologies, and research, development, demonstration, and deployment opportunities. [Online]. Available at: [www.epri.com](http://www.epri.com) [April 3, 2013].
196. Sokhansanj S and Hess JR, Biomass supply logistics and infrastructure. *Method Mol Biol* **581**:1–25 (2009).
197. Bravo ML, Naim M and Potter A, Key issues of the upstream segment of biofuels supply chain: A qualitative analysis. *Logist Res* **5**(1/2):21–31 (2012).
198. American Production and Inventory Control Society (APICS). Deciphering supply chain management and logistics. [Online]. Available at: <http://www.apics.org/news-landing-page/2012/04/25/deciphering-supply-chain-management-and-logistics> [April 3, 2013].
199. Yu H, Wanga Q, Ileleji K.E, Yu C, Luo Z and Gore KCJ, Design and analysis of geographic distribution of biomass power plant and satellite storages in China. Part 2: Road delivery. *Biomass Bioenerg* **46**:785–792 (2010).
200. Zhang F, Johnson DM and Johnson MA, Development of a simulation model of biomass supply chain for biofuel production. *Renew Energ* **44**:380–391 (2012).
201. Sokhansanj S, Turhollow A and Wilkerson E, *Integrated Biomass Supply and Logistics: ASABE Resource Magazine Engineering and Technology for a Sustainable World* (2008). AQ31
202. Judd JD, Sarin SC and Cundiff JS, Design, modeling, and analysis of a feedstock logistics system. *Bioresour Technol* **103**:209–18 (2012).
203. Clarke S and Preto F. *Biomass Densification for Energy Production*. OMAFRA Factsheet No 11-035 AGDEX 737/120 (2011).
204. Tumuluru JS, Wrigth CT, Hess JR and Kenney KL, A review of biomass densification system to develop uniform feedstock commodities for bioenergy application. *Biofuels Bioprod Bioref* **5**:683–707 (2011).
205. An H and Searcy SW, Economic and energy evaluation of a logistics system based on biomass modules. *Biomass Bioenerg* **46**:190–202 (2013).
206. Karunanithy C, Muthukumarappan K and Donepudi A, Moisture sorption characteristics of switchgrass and prairie cord grass. *Fuel* **103**:171–178 (2012).
207. Emery IR and Mosier NS, The impact of dry matter loss during herbaceous biomass storage on net greenhouse gas emissions from biofuels production. *Biomass Bioenerg* **39**:237–246 (2012).
208. Balat M, Global status of biomass energy use. *Energy Source A* **31**:1160–1173 (2009).
209. Balatinez JJ, The potential role of densification in biomass utilization. *Biomass Utilization* **67**:181–190 (1983).
210. Faaij A, Modern biomass conversion technologies. *Mitig Adapt Strat Glob Change* **11**:343–375 (2006).
211. Dijkman TJ and Benders RMJ, Comparison of renewable fuels based on their land use using energy densities. *Renew Sust Energ Rev* **14**:3148–3155 (2010).
212. Benoit G, Nguyen N, Maupu P and Vial E, Life cycle assessment of eucalyptus short rotation coppices for bioenergy production in southern France. *GCB Bioenergy* **5**:30–42 (2013).
213. Valente C, Spinelli R, Bengt Hillring BG. LCA of environmental and socio economic impacts related to wood energy production in alpine conditions: Valle di Fiemme (Italy). *J Clean Prod* **19**:1931–1938 (2011).
214. Valente C, Hillring BG and Solberg B, Bioenergy from mountain forest: A life cycle assessment of the Norwegian woody biomass supply chain. *Scand J Forest Res* **26**:429–436 (2011).



**Bruce Coulman**

Dr Bruce Coulman is a professor and head of the Plant Sciences department of the University of Saskatchewan. His present research focuses on the breeding and genetics of perennial forage grasses. Over his 37-year career, he has developed and released more than 20 new perennial grass and

legume cultivars.



### Ajay K. Dalai

Dr Ajay K. Dalai is a Tier 1 *Canada Research Chair in Bioenergy and Environmentally Friendly Chemical Processing* at the University of Saskatchewan. His research focus is the novel catalyst development for gas to liquid technologies, and biodiesel productions and applications. During his years as a researcher, Dr Dalai has published 250 research papers. He is a fellow of CIC, CAE and EIC.



### Camilo Perez Lee

Camilo Perez Lee is a Master's student in the field of Bioresource Engineering at McGill University. He is conducting research on harvest and post-harvest densification under the supervision of Dr Mark Lefsrud. Camilo's research focuses on the bioenergy supply chain planning and logistics from harvesting to biofuel. He holds a BSc in the field of Industrial Engineering.

AQ34

AQ32



### Emily Heaton

Dr Emily Heaton is \_\_\_\_\_ at Iowa State University. Her aim is to understand the growth and productivity of dedicated biomass crops, and how they can be managed to provide multiple ecosystem services. She specifically seeks to elucidate the reciprocal impact of environment on key physiological processes like photosynthesis, biomass accumulation, water use and nutrient cycling.



### Jaswinder Singh

Dr Jaswinder Singh researches enhancement of quality traits, stress tolerance and bioenergy capability of crops using genomic and biotechnological tools at McGill University. His current projects involve the assessment of diversity in cell walls and cell-wall synthesis genes in grasses; and the exploration of cultivated and wild germplasm for novel genes suitable for the next generation of cereal crops.

AQ33



### Donald L. Smith

Donald L. Smith – James McGill Professor at McGill University – conducts research on crop-microbe interactions, including biofuel feedstock crops. He has trained 60 graduate students, generated 8 patents, started a spin-off company, and commercialized technologies applied to millions of ha each year. He leads BioFuelNet Canada (\$12 million per year).



### Joann Whalen

Dr Joann Whalen is a soil ecologist at McGill University who investigates nutrient cycling in the soil-plant system. She is interested in biomass production systems that increase soil carbon sequestration and improve nutrient use efficiency, and evaluates system performance with life cycle assessment.



### Mark Lefsrud

Mark Lefsrud is an assistant professor of Bioresource Engineering at McGill University. He is currently conducting research on growing plants and other micro-organisms in controlled environments (greenhouses and growth chambers), and harvesting and handling biomass for biofuel.



### David B. Levin

David B. Levin leads a multidisciplinary research group focused on bioengineering for biofuels and bioproducts at the University of Manitoba. He co-leads a Genome-Canada-funded project on microbial genomics for biofuels and co-products from biorefining processes, and is Prairie Platform leader within BioFuelNet, a pan-Canadian research network funded by the Network Centres of Excellence program.

AQ35

**Peggy G. Lemaux**

Peggy G. Lemaux is \_\_\_\_\_ at the University of California, Berkeley, where efforts in the Lemaux laboratory have focused on engineering cereal crops, like sorghum, barley and wheat, to improve agronomic performance, nutritional profile, and suitability for biofuel use. Most recently, experimentation has turned to engineering tobacco to produce drop-in hydrocarbon biofuels in leaves.

**Sharon P. Shoemaker**

Dr Sharon P. Shoemaker is \_\_\_\_\_ at UC Davis. She is the author of several patents on yeasts to convert biomass to ethanol and bacterial strains to produce new forms of cellulose. She is researching the application of cellulases in biomass conversion, the integration of various unit operations in biomass conversion processes, and the development of new analytical methods for quantifying cellulose activity.

AQ36

**David B. Neale**

Dr David B. Neale is an adjunct geneticist and professor at the Department Plant Science at UC Davis. His research interests concern the genomics of forest trees, complex traits, QTL and association studies, population genetics and adaptation, and marker-based breeding. The focus has been on traits of practical value such as wood quality, growth, and disease resistance but will expand to genes determining adaptation and response to environmental stresses.

Uncorrected Proof

---

## QUERIES TO BE ANSWERED BY AUTHOR

**IMPORTANT NOTE: Please mark your corrections and answers to these queries directly onto the proof at the relevant place. Do NOT mark your corrections on this query sheet.**

---

### Queries from the Copyeditor:

- AQ1 refs 1-3 not cited in text. Please check
- AQ2 citations should be in numerical order. This is the first instance of ref 2 used. Please renumber previous references accordingly
- AQ3 citations should be in numerical order. This is the first instance of ref 3 used. Please renumber previous references accordingly
- AQ4 citations should be in numerical order. This is the first instance of ref 18-19 used. Please renumber previous references accordingly
- AQ5 Reference out of sequence. Please renumber - should be 54...
- AQ6 Ref out of sequence. Please renumber - should be 63
- AQ7 reference out of sequence. Please renumber. Should be 78
- AQ8 refs 99, 100, and 101 have not yet been cited.
- AQ9 ref out of sequence. Should be 170. Please renumber
- AQ10 ref out of sequence. Should be 171. Please renumber
- AQ11 ref out of sequence. Should be 172. Please renumber
- AQ12 ref out of sequence. Should be 173. Please renumber
- AQ13 all refs in this paragraph are out of sequence. Please renumber
- AQ14 please renumber references so that they are in sequence on first use
- AQ15 cannot locate citation 170 - please check
- AQ16 citations 172-176 missing
- AQ17 citations 180 and 181 missing
- AQ18 Please provide access date.
- AQ19 Please review url, shows page not found.
- AQ20 Please provide specific access date.
- AQ21 If Journal article, please provide name of journal and volume no. page range, and year of publication. If a book, please give name and location of publisher
- AQ22 Please provide date of meeting.
- AQ23 Please provide location of publisher.
- AQ24 Please provide location.
- AQ25 If Journal article, please provide name of journal and volume no. page range, and year of publication. If a book, please give name and location of publisher.
- AQ26 Please provide conference dates.
- AQ27 Please supply page range in(143-145, 147, 148, 150, 151, 153-155, 157, 158, 163, 165, 166, 167)
- AQ28 Please provide conference dates.
- AQ29 Please confirm jnl title. Showing as Geochim Cosmochim Ac
- AQ30 Please provide access date.
- AQ31 Please provide name of publisher and location.
- AQ32 please give current position
- AQ33 does this refer to turnover?
- AQ34 from which university?
- AQ35 please provide current position
- AQ36 please give current position

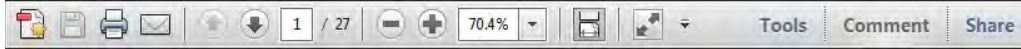


USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

Required software to e-annotate PDFs: [Adobe Acrobat Professional](#) or [Adobe Reader](#) (version 7.0 or above). (Note that this document uses screenshots from [Adobe Reader X](#))

The latest version of Acrobat Reader can be downloaded for free at: <http://get.adobe.com/uk/reader/>

Once you have Acrobat Reader open on your computer, click on the [Comment](#) tab at the right of the toolbar:



This will open up a panel down the right side of the document. The majority of tools you will use for annotating your proof will be in the [Annotations](#) section, pictured opposite. We've picked out some of these tools below:



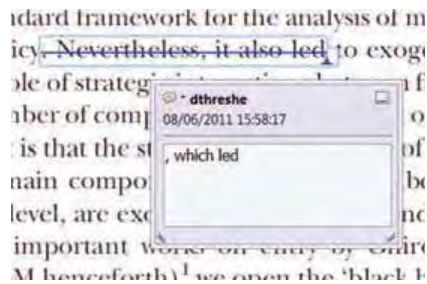
**1. Replace (Ins) Tool – for replacing text.**



Strikes a line through text and opens up a text box where replacement text can be entered.

**How to use it**

- Highlight a word or sentence.
- Click on the [Replace \(Ins\)](#) icon in the Annotations section.
- Type the replacement text into the blue box that appears.



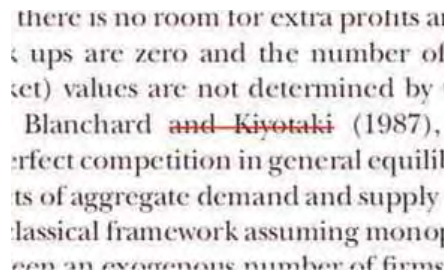
**2. Strikethrough (Del) Tool – for deleting text.**



Strikes a red line through text that is to be deleted.

**How to use it**

- Highlight a word or sentence.
- Click on the [Strikethrough \(Del\)](#) icon in the Annotations section.



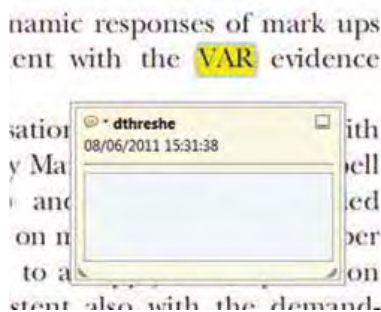
**3. Add note to text Tool – for highlighting a section to be changed to bold or italic.**



Highlights text in yellow and opens up a text box where comments can be entered.

**How to use it**

- Highlight the relevant section of text.
- Click on the [Add note to text](#) icon in the Annotations section.
- Type instruction on what should be changed regarding the text into the yellow box that appears.



**4. Add sticky note Tool – for making notes at specific points in the text.**



Marks a point in the proof where a comment needs to be highlighted.

**How to use it**


- Click on the [Add sticky note](#) icon in the Annotations section.
- Click at the point in the proof where the comment should be inserted.
- Type the comment into the yellow box that appears.





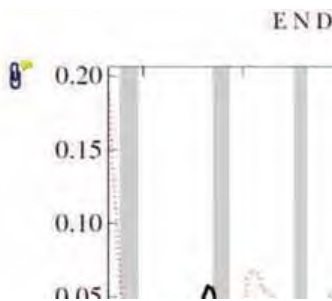
USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

**5. Attach File Tool – for inserting large amounts of text or replacement figures.**


 Inserts an icon linking to the attached file in the appropriate place in the text.

**How to use it**

- Click on the [Attach File](#) icon in the Annotations section.
- Click on the proof to where you'd like the attached file to be linked.
- Select the file to be attached from your computer or network.
- Select the colour and type of icon that will appear in the proof. Click OK.



**6. Add stamp Tool – for approving a proof if no corrections are required.**

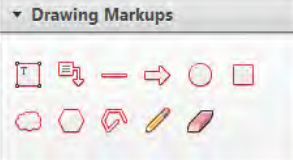
 Inserts a selected stamp onto an appropriate place in the proof.

**How to use it**

- Click on the [Add stamp](#) icon in the Annotations section.
- Select the stamp you want to use. (The [Approved](#) stamp is usually available directly in the menu that appears).
- Click on the proof where you'd like the stamp to appear. (Where a proof is to be approved as it is, this would normally be on the first page).

of the business cycle, starting with the  
 on perfect competition, constant ret  
 production. In this minimum cost  
 es  
 he  
 s  
 otaki (1987), has introduced produc  
 general equilibrium models with nomin  
 ed and unbalanced. Most of this litera

**APPROVED**

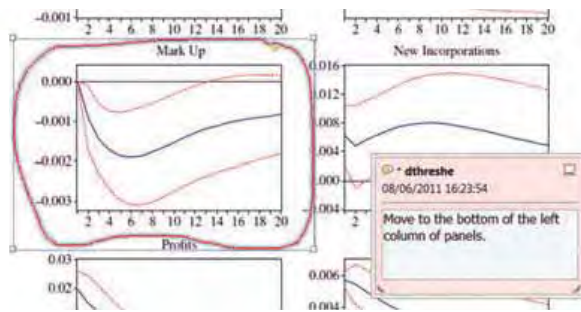


**7. Drawing Markups Tools – for drawing shapes, lines and freeform annotations on proofs and commenting on these marks.**

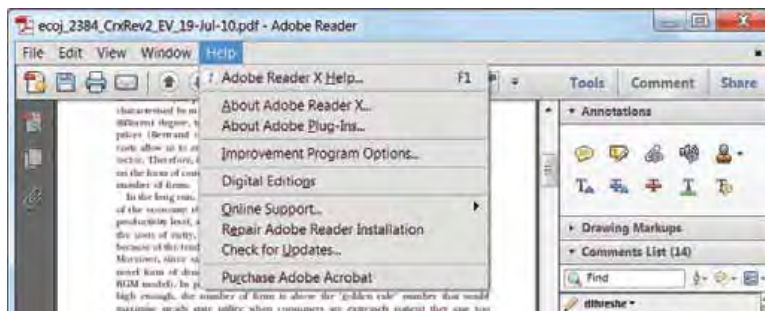
Allows shapes, lines and freeform annotations to be drawn on proofs and for comment to be made on these marks..

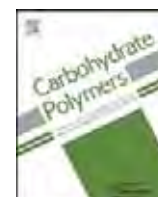
**How to use it**

- Click on one of the shapes in the [Drawing Markups](#) section.
- Click on the proof at the relevant point and draw the selected shape with the cursor.
- To add a comment to the drawn shape, move the cursor over the shape until an arrowhead appears.
- Double click on the shape and type any text in the red box that appears.



For further information on how to annotate proofs, click on the [Help](#) menu to reveal a list of further options:





# Characterization of ionic liquid pretreated aspen wood using semi-quantitative methods for ethanol production

Amir Goshadrou<sup>a,b</sup>, Keikhosro Karimi<sup>b,c,\*</sup>, Mark Lefsrud<sup>a</sup>

<sup>a</sup> Department of Bioresource Engineering, McGill University, 2111 Lakeshore Road, Ste. Anne de Bellevue, Montreal, Quebec H9X 3V9, Canada

<sup>b</sup> Department of Chemical Engineering, Isfahan University of Technology, Isfahan 84156-83111, Iran

<sup>c</sup> Industrial Biotechnology Group, Institute of Biotechnology and Bioengineering, Isfahan University of Technology, Isfahan 84156-83111, Iran

## ARTICLE INFO

### Article history:

Received 20 December 2012

Received in revised form 10 April 2013

Accepted 11 April 2013

Available online xxx

### Keywords:

Lignocellulose

Pretreatment

Simons' stain

Accessibility

Crystallinity

Bioethanol

## ABSTRACT

Aspen wood (*Populus tremula*) was pretreated with ionic liquid 1-ethyl-3-methylimidazolium acetate ([EMIM]OAc) and dilute sulfuric acid for improvement of ethanol production. The ionic liquid pretreatment included wood dissolution at 120 °C and 5% solid loading for 1, 3, and 5 h followed by regeneration using water as an anti-solvent. More than 95% enzymatic digestibility was achieved for the ionic liquid treated wood, while the yield from the untreated wood was only 5.3%. Furthermore, over 81% of the maximum theoretical ethanol yield was attained after 24 h fermentation of the ionic liquid treated wood, whereas the yields were only 5.3% and 42.1% for the untreated and dilute acid treated materials, respectively. A side-by-side comparative analysis of the pretreated materials using semi-quantitative techniques (e.g., Simons' staining and enzyme adsorption) revealed that the ionic liquid treatment was much more successful in increasing the cellulose accessibility to cellulases and decreasing the lignin content.

© 2013 Elsevier Ltd. All rights reserved.

## 1. Introduction

Lignocellulose is the primary component in plant cell walls and is the most abundant sources of biomass in the world. It has a high potential to provide an inexpensive and renewable feedstock for the future production of bioethanol (Demirbas, 2011; Nigam & Singh, 2011; Sanchez & Cardona, 2008). Lignocelluloses have a complex and rigid structure, mainly composed of carbohydrate polymers (cellulose and hemicelluloses), lignin and other minor components. Cellulose is tightly bound to lignin and hemicellulose through hydrogen and covalent bonding in the matrix structure of the cell wall (Seeta Laxman & Lachke, 2009; Taherzadeh & Karimi, 2008a). Saccharification, the process of breaking down cellulose into simple sugars, is an essential stage to allow the conversion to bioethanol. However, the recalcitrant nature of lignocellulose limits its chemical and biological degradation, and thus hinders the hydrolysis process (Taherzadeh & Karimi, 2007, 2008b).

The major obstacle of hydrolysis of lignocellulose is not only due to the protective nature of lignin and hemicelluloses, but also the intrinsic physical and morphological characteristics of the heterogeneous biomass such as cellulose fibers crystallinity and surface

area, which govern the hydrolysis efficacy. The cellulose crystallinity is defined as the ratio of the crystalline to the amorphous portions that influences the rate of hydrolysis, given that the loosely organized amorphous cellulose is more amiable to enzymatic (cellulase) conversion than the closely packed crystalline cellulose. The available surface area (exterior and interior), porosity, and range of pore sizes play significant roles in the rate and extent of enzymatic hydrolysis, by providing active sites for cellulase adsorption and facilitating enzyme diffusion into the pores (Alvira, Tomas-Pejo, Ballesteros, & Negro, 2010; Chandra, Ewanick, Hsieh, & Saddler, 2008; Chandra, Esteghlalian, & Saddler, 2008).

Pretreatment is a key step to overcome the lignocelluloses recalcitrance and make them more susceptible for hydrolysis by cellulases. Several methods have been discussed in the literatures including physical (e.g., mechanical comminution and pyrolysis), chemical (e.g., acid hydrolysis and organosolv process), physicochemical (e.g., steam explosion and AFEX), and biological pretreatments (e.g., fungi and actinomycetes). One of the primary challenges is the development of a cost-effective pretreatment technology, while remaining environmental and energy efficient. Some pretreatment techniques such as acid hydrolysis are hazardous, corrosive, and need neutralization due to the formation of inhibitory compounds, whereas others like steam explosion and AFEX are capital intensive (Chandra et al., 2007; Gupta, Khasa, & Kuhad, 2011; Taherzadeh & Karimi, 2008b; Yang & Wyman, 2008).

Certain ionic liquids have demonstrated a promising ability for efficient dissolution of biomass and its cellulose regeneration upon

\* Corresponding author at: Department of Chemical Engineering, Isfahan University of Technology, Isfahan 84156-83111, Iran. Tel.: +98 3113915623; fax: +98 3113912677.

E-mail address: [karimi@cc.iut.ac.ir](mailto:karimi@cc.iut.ac.ir) (K. Karimi).

dilution with an anti-solvent (Diedericks, van Rensburg, Garcia-Aparicio, & Gorgens, 2012; Geng & Henderson, 2012; Li et al., 2010; Moniruzzaman & Ono, 2012; Tan & Lee, 2012; Uju et al., 2012). These ionic solvents are liquid at or near room temperature, and typically are composed of large organic cations and small inorganic anions. Some ionic liquids are green solvents, and have several attractive properties, such as excellent chemical and thermal stability, non-flammability, and very low to negligible vapor pressure (Zhu et al., 2006). Pretreatment with ionic liquids can be applied at moderate conditions with the possibility of almost complete recovery of the solvent for subsequent reuse (Zhao et al., 2009). Among the ionic liquids, 1-alkyl-3-methylimidazolium salts such as 1-ethyl-3-methylimidazolium acetate ([EMIM]OAc) and 1-butyl-3-methylimidazolium ([BMIM]Cl) are found to be very efficient solvents for cellulose dissolution (Auxenfans et al., 2012; Mora-Pale, Meli, Doherty, Linhardt, & Dordick, 2011). It has been reported that out of six different ionic liquids, [EMIM]OAc significantly enhanced sugarcane bagasse enzymatic saccharification yield over 90% (Sant'Ana da Silva, Lee, Endo, & Bon, 2011). Furthermore, [EMIM]OAc does not have a reactive side group, and is non-toxic, non-corrosive, and biodegradable (El Seoud, Koschella, Fidale, Dorn, & Heinze, 2007; Zhu et al., 2006). Despite the high potential of ionic liquids, there is limited knowledge dealing with effects of ionic liquids on features of lignocellulosic substrates that promote their hydrolytic susceptibility.

In order to track the changes in lignocellulose properties due to the pretreatment, several quantitative methods such as nitrogen adsorption and solute exclusion have been used to measure the substrate surface area and accessibility to cellulases (Chandra, Ewanick, et al., 2008; Chandra, Esteghlalian, et al., 2008). There are also a few reports on application of some more quick methods such as Simons' stain technique and water retention value. The Simons' stain method was originally developed as a staining method for evaluation of the mechanical damage of pulp fibers during the beating process. However, it has shown levels of success in estimating a given pretreated substrate's susceptibility to enzymatic hydrolysis, and thus the effectiveness of the pretreatment (Chandra, Ewanick, et al., 2008; Chandra, Esteghlalian, et al., 2008).

This research evaluated the effectiveness of ionic liquid [EMIM]OAc pretreatment of aspen wood compared to dilute sulfuric acid by assessing the changes in physicochemical and structural properties of lignocellulosic substrate. The research focused on the use of rapid semi-quantitative methods to gain a better understanding of the key factors affecting enzymatic digestibility and subsequent bioconversion to ethanol.

## 2. Materials and methods

### 2.1. Raw material

An aspen wood (*Populus tremula*) log was harvested on summer 2011 from a local forest (Mazandaran, Iran, 36°28'N, 51°54'E, 480 m). The sample was debarked, cut into small pieces, and dried at room temperature. The wood chips were milled and screened to achieve a particle size in the range of 177–840  $\mu\text{m}$ . Moisture content of the wood was determined by drying samples in duplicates at 105 °C for 24 h.

### 2.2. Pretreatments

#### 2.2.1. Ionic liquid pretreatment

Ionic liquid, 1-ethyl-3-methylimidazolium acetate ([EMIM]OAc), produced commercially by BASF (Ludwigshafen, Germany), was purchased from Sigma–Aldrich (St. Louis, MO, USA). An amount of 1 g of the aspen wood (dry basis) was thoroughly

mixed with 19 g of the ionic liquid in 50 ml pressure glass bottles at 120 °C for increasing time periods (1, 3, and 5 h). After dissolution of the aspen wood in the ionic liquid for different pretreatment times, an equal volume of deionized water was added into the wood solutions to stop the treatment and precipitate the cellulose under vigorously stirring condition. The regenerated materials were filtered (Whatman Grade No. 41) under vacuum pressure and washed with boiling water until a clear filtrate was appeared. All pretreated materials were then oven dried at 40 °C for 48 h, weighed, and stored in sealed bags at 4 °C until use.

#### 2.2.2. Dilute sulfuric acid pretreatment

The aspen wood was presoaked in dilute sulfuric acid solution (1.2%, w/w) at a solid loading of 5% (w/w) and room temperature for 4 h. The slurry was heated up (10 °C/min) to reach the final temperature of 120 °C and then isothermally kept at 120 °C for 60 min (Pinto & Kamden, 1996). After pretreatment, the mixture was filtered under vacuum pressure, and the recovered materials were washed by distilled water to reach a pH of 6.0. The pretreated materials were oven dried at 40 °C for 48 h, weighed, and kept in sealed bags at 4 °C until use.

### 2.3. Materials characterization

#### 2.3.1. Composition analysis

Polymeric carbohydrates, acid soluble lignin, and acid insoluble lignin contents of the untreated and pretreated wood were determined by the method presented by the National Renewable Energy Laboratory (Sluiter et al., 2008).

#### 2.3.2. Simons' stain method

Accessibility of the woody materials to the cellulases was assessed by evaluating the pore size distribution of the samples using Simons' stain method. This method is based on the competitive adsorption of two dyes, Direct Blue 1 (DB) and Direct Orange 15 (DO), in an aqueous medium, and in this study, a modified version of Simons' stain technique was used (Chandra, Ewanick, et al., 2008; Chandra, Esteghlalian, et al., 2008). DB (Pontamine Fast Sky Blue 6BX) and DO (Pontamine Fast Orange 6RN) dyes were provided by Pylam Products (Garden City, NY). The low-molecular-weight (LMW) fraction of the DO was filtered through a 100k molecular mass cut-off membrane, and the retentate was collected to make the DO stock solution. Approximately, 100 mg of the treated or untreated wood was weighed into 15 ml test tubes, and 1.0 ml of phosphate buffered saline solution (pH 6, 0.3 M  $\text{PO}_4$ , 1.4 mM NaCl) was added to each tube. Different amounts of DB and DO stock solutions (10 mg/ml), ranging from 0.25 to 2 ml, were added to the tubes and diluted up to 10 ml with distilled water, resulting in a ratio of 1:1 DO and DB at increasing concentrations. All the tubes were incubated at 70 °C and 150 rpm for 15 h and then centrifuged at 7000  $\times$  g for 5 min. The amounts of DO and DB adsorbed by the wood was calculated using the difference between the initial and final dye concentration in the supernatant solution measured by UV-Visible spectrophotometer at wavelengths of 624 and 455 nm, respectively (Chandra et al., 2009; Esteghlalian, Bilodeau, Mansfield, & Saddler, 2001). The spectrophotometer was calibrated by preparing standard curves of each dye and measuring the absorbance at 455 and 624 nm.

#### 2.3.3. Water retention value

Water retention value (WRV) is the ability of a substrate to keep water molecules in the cell wall pores and is used to represent the total volume of the pores. The WRV measurement is based on subjecting a water-saturated sample to a centrifugal force, and the WRV is defined as the amount of retained water per unit weight of dry material. Approximately 1 g of the substrate was mixed with

**Table 1**  
Chemical composition (%) of untreated and pretreated aspen wood.<sup>a</sup>

Component	Untreated aspen wood	Dilute acid pretreated aspen wood	[EMIM]OAc pretreated aspen wood (time)		
			1 h	3 h	5 h
Glucan	49.0	56.1	59.3	60.4	63.9
Xylan	14.9	7.5	9.9	9.2	8.6
Mannan	2.0	1.9	1.7	1.5	1.7
Galactan	0.5	ND <sup>b</sup>	ND	ND	ND
Arabinan	0.8	ND	ND	ND	ND
Acid soluble lignin	1.0	0.8	1.1	1.0	1.0
Acid insoluble lignin	24.6	25.3	20.8	19.2	16.4

<sup>a</sup> The data are average values of the two replications.

<sup>b</sup> Not detectable.

deionized water at 150 rpm for 60 min. The mixture was filtered using a nylon membrane (45 μm), and the filter cake was transferred into a small bag made of nonwoven materials and soaked in deionized water at room temperature for 2 h. The bag was wrapped, placed into a centrifuge tube with support to make space for water accumulation, and centrifuged at 3000 × g for 15 min. The substrate was collected and weighed before and after drying at 105 °C for 24 h. The WRV was calculated as follows (Luo & Zhu, 2011; Program, 2000):

$$\text{WRV (g/g dry material)} = \frac{W_{\text{wet}} - W_{\text{dry}}}{W_{\text{dry}}} \quad (1)$$

where  $W_{\text{wet}}$  and  $W_{\text{dry}}$  are respectively the wet and oven dry mass of the material.

### 2.3.4. Substrate accessibility to cellulase

The maximum extent of enzyme adsorption, an indication of the available surface area for cellulase binding, was measured according to the method presented by Kumar and Wyman (2009) with modifications as follows. Different amounts of enzyme were dissolved in 5 ml sodium citrate buffer (50 mM, pH 4.8 ± 0.1) and added to the pretreated and untreated substrates to make suspensions with a final dry solid content of 2% (w/v). The mixtures were continuously stirred and preserved at 4 °C to avoid changes in substrate properties due to the possible hydrolysis. After 90 min, the supernatants were collected and centrifuged at 5000 × g for 15 min. The free protein content of the supernatants was measured by Bio-Rad protein assay (Richmond, CA, USA) with bovine serum albumin as a protein standard. The bonded protein was indirectly calculated by subtracting the measured free protein from the total protein added. The Langmuir isotherm model was used to describe the equilibrium adsorption behavior according to the following equation:

$$E_a = \frac{E_m K_d E_f}{1 + K_d E_f} \quad (2)$$

where  $E_a$  is the concentration of bound protein (mg protein/g substrate),  $E_m$  is the maximum adsorption capacity based on complete monolayer coverage assumption (mg protein/g substrate),  $E_f$  is the free protein concentration in the supernatant (mg protein/ml), and  $K_d$  is the binding constant (ml/mg protein). The Langmuir parameters ( $E_m$  and  $K_d$ ) were determined by non-linear regression of the experimental adsorption data based on the least-square method using Curve Fitting Toolbox of MATLAB software (Version R2011a; Math Works Inc., MA, USA).

### 2.3.5. FTIR spectroscopy

Fourier transform infrared (FTIR) spectroscopy technique was applied to examine changes in structure of the woody materials caused by the pretreatments. The FTIR system (Tensor 27 FT-IR Spectrometer, Bruker, Germany) was equipped with a universal attenuated total reflection sampling accessory, and deuterated

triglycine sulfate detector. The spectra were acquired at 4 cm<sup>-1</sup> resolution with an accumulation of 60 scans per sample, and recorded in the range of 600–4000 cm<sup>-1</sup>.

### 2.3.6. SEM

Scanning electron microscopy (SEM) analysis was performed to obtain detailed microscopic view of morphologic changes occurred by the pretreatments. Prior to the study, the untreated and pretreated samples were coated with gold (BAL-TEC SCD 005), and then images were recorded with an acceleration voltage of 15 kV (Philips XL30).

### 2.3.7. Enzymatic digestibility of the substrate

Two commercial enzymes, cellulase from *Trichoderma reesei* (Celluclast 1.5 L, 90 FPU/ml, Sigma) and cellobiase from *Aspergillus niger* (Novozyme 188, 744 CBU/ml, Sigma), were used for hydrolysis experiments. Enzyme activities were determined according to standard IUPAC procedures (Ghose, 1987). Enzymatic digestibility of untreated and pretreated materials was conducted in 50 mM sodium citrate buffer (pH 4.8 ± 0.1) at 3% (w/v) solids concentration, 50 ± 1 °C, and 150 rpm for 72 h. Enzymes were loaded at 20 FPU cellulase and 30 CBU β-glucosidase per gram glucan and the suspension was supplemented with 0.5 g/l sodium azide to restrict any microbial growth during the hydrolysis (Goshadrou, Karimi, & Taherzadeh, 2011). Samples were taken periodically over 72 h, immersed in boiling water for 5 min to deactivate the enzymes, and stored at -20 °C before sugar analysis. Pure cellulose (Avicel PH-101, Sigma) as a control was run alongside the samples in the hydrolysis experiments. The amount of glucose released during digestibility experiments was determined enzymatically using glucose oxidase-peroxidase assay kit (Sigma). The substrate enzymatic digestibility was defined as:

$$\text{Enzymatic digestibility (\%)} = \frac{\text{Produced glucose (g/l)} \times 100}{\text{Substrate concentration (g/l)} \times F \times 1.111} \quad (3)$$

where  $F$  is the biomass glucan fraction (Table 1), and the constant 1.111 is used for conversion of glucan to glucose.

## 2.4. Microorganism and media

A flocculating strain of *Saccharomyces cerevisiae* CCUG 53310 (Culture Collection University of Göteborg, Göteborg, Sweden) was used for ethanol production. The yeast was cultivated on agar plates contained (g/l): glucose, 20; peptone, 20; and yeast extract 10. After two days of incubation at 30 ± 0.5 °C, the plates were stored at 4 °C until use.

## 2.5. Inoculum preparation

The biomass for fermentation was prepared in 250 ml cotton-plugged Erlenmeyer flasks containing 50 g/l glucose and



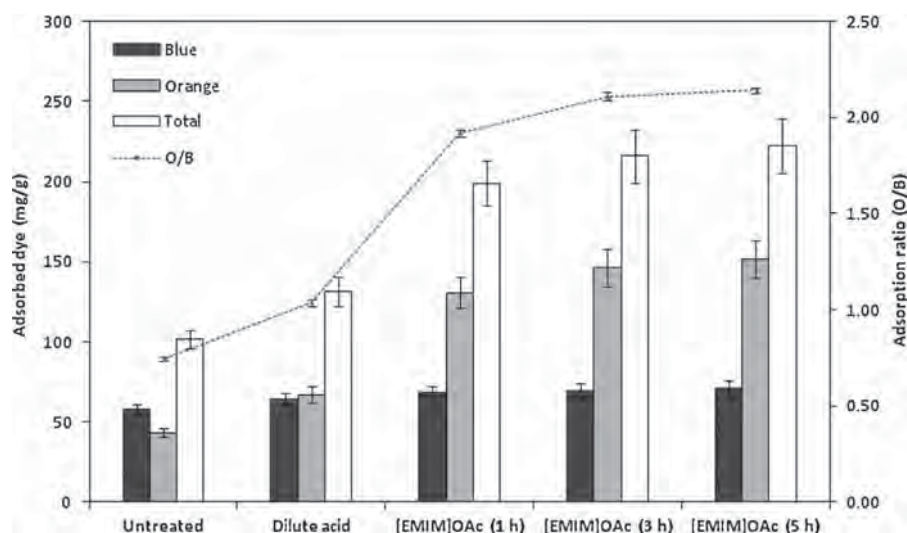


Fig. 1. Dyes adsorption during Simons' stain analysis for untreated and pretreated aspen wood.

necessary nutrients including (g/l): yeast extract, 5;  $(\text{NH}_4)_2\text{SO}_4$ , 7.5;  $\text{K}_2\text{HPO}_4$ , 3.5;  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 0.75;  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ , 1; in distilled water (Goshadrou, Karimi, & Taherzadeh, 2013). The pH was adjusted to  $5.0 \pm 0.1$  using either 2.5 M sodium hydroxide or 0.5 M sulfuric acid. The media was autoclaved at  $121^\circ\text{C}$  for 20 min, inoculated by adding a single colony of the yeast and then incubated for 30 h at  $30 \pm 0.5^\circ\text{C}$  and 150 rpm. The produced biomass was then aseptically harvested by centrifugation at  $3400 \times g$  for 15 min (D-78532 Tuttlingen, Hettrich, Germany), and the supernatant was removed.

## 2.6. Bioconversion to ethanol

The sugar solutions following 30 h enzymatic hydrolysis of the untreated and pretreated aspen wood (without sodium azide addition) were supplemented with necessary nutrients (g/l): yeast extract, 5.0,  $(\text{NH}_4)_2\text{SO}_4$ , 7.5,  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 0.75,  $\text{K}_2\text{HPO}_4$ , 3.5, and  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ , 1.0. The media were autoclaved at  $121^\circ\text{C}$  for 20 min, cooled to room temperature, and aseptically inoculated with 5 g/l yeast biomass (based on the dry mass). The samples were incubated at  $35 \pm 0.5^\circ\text{C}$  and 150 rpm for 30 h. All fermentations were performed in 115 ml serum bottles sealed with rubber stoppers. The bottles were purged with pure nitrogen at the beginning of the cultivation and vented with syringe needles to release the  $\text{CO}_2$  during fermentation. Ethanol yield was calculated as a percentage of maximum theoretical yield according to the following equation:

$$\text{Ethanol yield (\%)} = \frac{\text{Produced ethanol (g/l)} \times 100}{\text{Substrate concentration (g/l)} \times F \times 1.111 \times 0.51} \quad (4)$$

The factor 0.51 was applied to consider the theoretical conversion of glucose to ethanol.

## 2.7. Analytical methods

The liquid samples from fermentation were analyzed using an HPLC system equipped with UV-Vis and refractive index detectors (Jasco International Co., Tokyo, Japan). The ethanol concentration was determined after separation on an Aminex HPX-87H ion-exchange column (Bio-Rad, Richmond, CA, USA) at  $60^\circ\text{C}$  with 0.6 ml/min eluent of 5 mM  $\text{H}_2\text{SO}_4$ . The concentration of the sugars in composition analysis (Section 2.3.1) was analyzed using an

ion-exchange Aminex HPX-87P column (Bio-Rad, Richmond, CA, USA) at  $85^\circ\text{C}$  with 0.6 ml/min ultra-pure water as eluent.

All the analyses and experiments were performed in duplicate, and the reported data are average values of the two replications.

## 3. Results

### 3.1. Effect of pretreatment on aspen wood composition

The main chemical composition of control (untreated), dilute sulfuric acid, and ionic liquid pretreated aspen woods were analyzed, and the results are presented in Table 1. The pretreatments resulted in materials with different cellulose and hemicellulose contents. The untreated wood contained 49.0% glucan which increased to 56.1% and 59.3–63.9% after dilute acid and ionic liquid pretreatments, respectively. All the pretreatments removed the xylan and other polysaccharides and this was more pronounced after using the dilute acid pretreatment. The acid soluble lignin content of the wood was low ( $1.0 \pm 0.1\%$ ) and remained unchanged after the ionic liquid pretreatments, while it decreased to 0.8% by the dilute acid pretreatment. The acid insoluble parts of lignin were reduced from 24.6% to 16.4–20.8% during ionic liquid treatments, while it increased slightly after dilute acid treatment. Increasing ionic liquid pretreatment time highly impacted the lignin removal yield.

### 3.2. Effect of pretreatment on surface area and pore sizes

The two-color Simons' stain experiments were carried out on the untreated and pretreated woody materials to quantify changes in the biomass porosity and pores sizes as a result of the pretreatments. The results (Fig. 1) indicated that the total adsorbed dyes increased from 101 to about 130 mg/g using dilute acid treatment, whereas the extent of adsorption was significantly higher for biomass treated with ionic liquid, ranging from 199 to 222 mg/g for 1 and 5 h treated materials, respectively. Adsorption ratio (DO/DB) is defined as the maximum adsorbed DO by the substrate over that of DB. The adsorption ratio was measured as 0.75 for untreated aspen wood, while it increased to 1.92, 2.11, and 2.14 after 1, 3, and 5 h pretreatment with the ionic liquid, respectively. No significant difference between 3 and 5 h pretreatment times was detected. Among the pretreated woods, the lowest ratio of 1.04 was observed for the biomass treated with dilute acid.



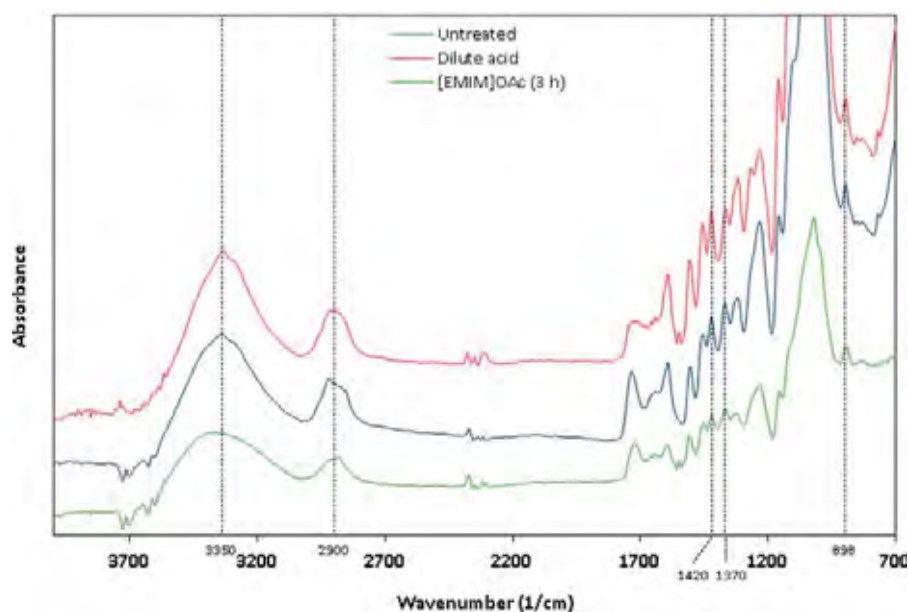


Fig. 2. FTIR spectra of untreated and pretreated aspen wood.

### 3.3. Effect of pretreatment on pore volumes

The swelling ability of the biomass before and after pretreatments was characterized by the water retention value as it mainly reflects the pores volume of a substrate. For determination of WRV values, the substrate was soaked in deionized water, subjected to centrifugal force, oven-dried, and then the ratio of absorbed water to dry fibers was calculated. As shown in Table 2, WRV of the untreated aspen wood was 0.71 g/g dry material, which significantly increased up to 1.87 g/g dry material after the ionic liquid treatment. The extent of water swelling capacity was less affected by ionic liquid treatment time. Unexpectedly, the acid pretreated wood exhibited the least ability to absorb water molecules even in comparison with untreated aspen wood.

### 3.4. Effect of pretreatment on accessible surface area

Langmuir adsorption isotherm was applied to describe cellulase binding to the untreated and pretreated substrates, and the

Table 2

Water retention values (WRV) of untreated and pretreated aspen wood.

Substrate	WRV <sup>a</sup> (g/g dry material)
Dilute acid treated	0.67
[EMIM]OAc treated (1 h)	1.81
[EMIM]OAc treated (3 h)	1.85
[EMIM]OAc treated (5 h)	1.87
Untreated	0.71

<sup>a</sup> The data are average values of the two replications.

Table 3

Langmuir isotherm parameters for cellulase binding to untreated and pretreated materials.

Substrate	Langmuir isotherm parameters		$S = E_{\max} \times K_d$ (ml/g substrate)	$R^2$
	$E_{\max}$ (mg/g substrate)	$K_d$ (ml/mg)		
Aspen wood				
Untreated	101.6	1.01	102.6	0.958
Dilute acid treated	121.6	1.12	136.2	0.967
[EMIM]OAc treated (1 h)	169.9	1.37	232.8	0.964
[EMIM]OAc treated (3 h)	175.4	1.38	242.0	0.961
[EMIM]OAc treated (5 h)	176.8	1.35	238.7	0.963
Pure cellulose	88.9	0.34	30.2	0.966

maximum extent of adsorbed protein was considered proportional to the surface area available for the cellulase enzyme. Adsorption capacity ( $E_{\max}$ ), affinity parameter ( $K_d$ ), and binding strength ( $S$ ) are listed in Table 3, along with the determination coefficient ( $R^2$ ) in the range of 0.958–0.967. As shown in Table 3, the adsorption capacity of the aspen wood was 101.6 mg/g, which was increased to 121.6–176.8 mg/g after the pretreatment. Among the pretreated materials, the solids remained after the dilute acid treatment showed the lowest adsorption capacity (121.6 mg/g solids), whereas 5 h ionic liquid treated wood was the most accessible substrate to the cellulase. However, no significant improvement in cellulase adsorption behavior was observed between samples treated with ionic liquid for 3 and 5 h. The binding affinity of cellulase to the samples increased from 1.01 to 1.12–1.38 ml/mg after pretreatment using dilute acid and ionic liquid, respectively.

### 3.5. Effect of pretreatment on substrate crystallinity

The bands for untreated and representative treated materials were observed over the range of 600–4000  $\text{cm}^{-1}$ . The absorbance spectra and characteristic of the bands are presented in Fig. 2 and Table 4, respectively. The band intensities associated with lignin at 1240  $\text{cm}^{-1}$  (C–O vibration in syringyl and guaiacyl condensed lignin), 1460  $\text{cm}^{-1}$  (asymmetric bending in C–H<sub>3</sub> in lignin), 1510  $\text{cm}^{-1}$  (aromatic skeleton C=C stretching in lignin), and 1590/1627  $\text{cm}^{-1}$  (C=C stretching vibration in lignin) significantly decreased for the 3 h ionic liquid treated materials compared to untreated aspen wood, whereas the dilute acid treated solid showed higher lignin content than the other woody materials.

**Table 4**

Characteristic of bands in FTIR spectra of untreated and pretreated materials.

Wavenumber (cm <sup>-1</sup> )	Functional group	Assignment	Intensity		
			Untreated	Dilute acid	[EMIM]OAc (3 h)
1055	C–O stretching	Cellulose and hemicellulose	0.680	0.938	0.284
1158	C–O–C asymmetric stretching	Cellulose	0.300	0.414	0.143
1240	C–O stretching	Hemicellulose and lignin	0.294	0.340	0.151
1460	Asymmetric bending in C–H <sub>3</sub>	Lignin	0.176	0.263	0.110
1510	C=C stretching of the aromatic ring	Lignin	0.140	0.226	0.098
1590	C=C stretching	Lignin	0.147	0.211	0.088
1627	C=C stretching	Lignin	0.115	0.165	0.076
1730	C=O stretching of acetyl or carboxylic acid	Hemicellulose and lignin	0.136	0.161	0.086
3350	O–H stretching	Cellulose	0.175	0.231	0.100
Aspen wood		LOI (A1420/A898) <sup>a</sup>		TCI (A1375/A2900) <sup>a</sup>	
Dilute acid treated		1.88		0.98	
[EMIM]OAc treated (3 h)		1.12		0.82	
Untreated		1.66		1.05	

<sup>a</sup> The crystallinity indices were determined following the method of Nelson and O'Connor (1964).

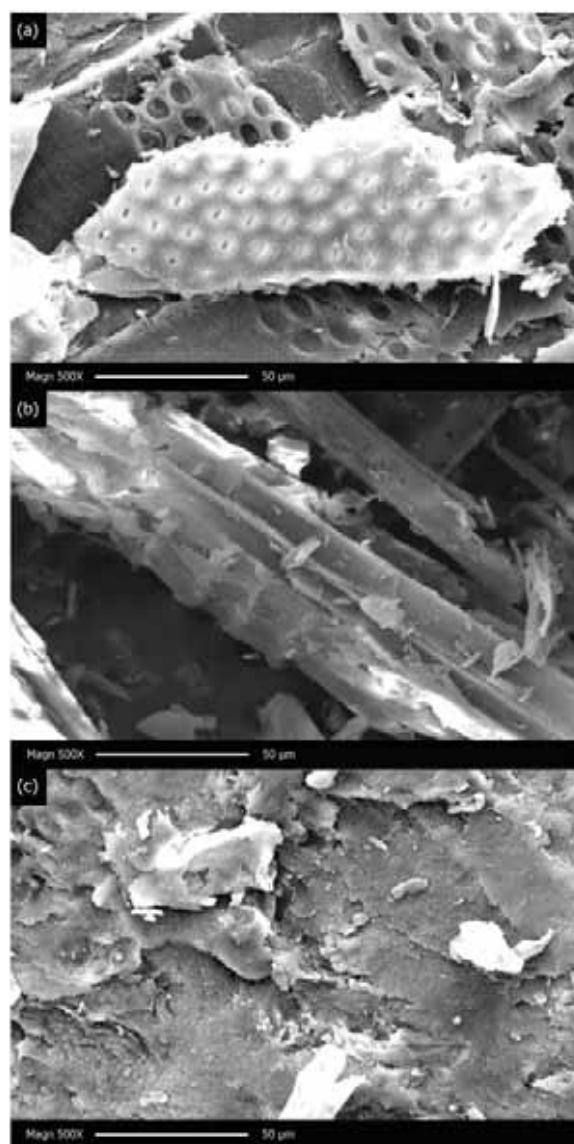
Furthermore, the spectral ratios A1420/A898 and A1370/A2900, which are respectively referred as crystallinity index or lateral order index (LOI) and total crystallinity index (TCI), were calculated (Carrillo, Colom, Sunol, & Saurina, 2004; Nelson & O'Connor, 1964; Uju et al., 2012). Both LOI and TCI indices were efficiently impacted by the ionic liquid pretreatment and dropped from 1.66 and 1.05 to 1.12 and 0.82, respectively. However, the total crystallinity index was less impacted by the dilute acid pretreatment and reduced to 0.98. Besides, LOI of the acid treated materials (1.88) showed that the pretreatment resulted in slightly higher crystallinity than the untreated biomass. The bands related to intra- and intermolecular O–H stretching of cellulose hydroxyl group can be detected in the range of 3000–3600 cm<sup>-1</sup> (Oh et al., 2005). As it is evident from Fig. 2, the spectrum specifically at 3350 cm<sup>-1</sup> was weakened, broadened, and shifted to a higher wavenumber for the ionic liquid treated sample indicated the alteration of hydrogen bonds network in the cellulose structure and lower crystallinity. The same trend was observed for the typical peaks of cellulose and hemicellulose at 1730 cm<sup>-1</sup>. The results indicated that the ionic liquid treatment was more successful in modifying the lignocellulose structure.

### 3.6. Effect of pretreatment on substrate morphology

Scanning electron microscopy analysis was conducted to qualitatively assess the surface morphology of the untreated and pretreated materials. As shown in Fig. 3, the dilute acid pretreatment resulted in a modified and destructed morphological structure compared with the untreated wood. However, it was apparent that the major microfibrils in the cell wall structure was not highly affected by the acid treatment. Conversely, a different supramolecular structure was observed for wood treated with [EMIM]OAc. The fibrillar pattern was completely disrupted and a spongier structure, with more conglomerate texture, was formed by the pretreatment. Also, the surface of the ionic liquid treated wood appeared to be disordered.

### 3.7. Effect of pretreatment on substrate enzymatic digestibility

The untreated and different pretreated substrates along with pure cellulose (Avicel PH-101) as a reference were subjected to enzymatic hydrolysis, and a summary of the results are presented in Table 5. Without pretreatment, no appreciable amount of glucose was produced from the wood even after 72 h of hydrolysis. However, digestibility of the lignocellulose was efficiently improved by both dilute acid and ionic liquid pretreatments. The cellulose conversion progressed quickly during the first 24 h, slowed until 48 h,



**Fig. 3.** Scanning electron micrographs of (a) untreated, (b) dilute acid, and (c) [EMIM]OAc (3 h) pretreated aspen wood.

**Table 5**  
Enzymatic digestibility of untreated and pretreated materials.<sup>a</sup>

Substrate	Enzymatic digestibility <sup>b</sup> (%)			
	6 h	24 h	48 h	72 h
Aspen wood				
Untreated	4.4 ± 0.1	5.3 ± 0.2	5.2 ± 0.2	5.3 ± 0.2
Dilute acid treated	27.2 ± 0.8	52.5 ± 0.6	56.1 ± 1.1	59.7 ± 1.9
[EMIM]OAc treated (1 h)	40.2 ± 0.5	78.3 ± 1.0	82.5 ± 2.0	82.1 ± 1.7
[EMIM]OAc treated (3 h)	44.4 ± 0.7	86.0 ± 1.1	89.8 ± 1.6	91.3 ± 1.3
[EMIM]OAc treated (5 h)	49.9 ± 1.1	87.3 ± 0.9	94.3 ± 1.5	95.4 ± 1.0
Pure cellulose	26.4 ± 0.3	45.2 ± 1.1	51.2 ± 1.4	53.3 ± 0.8

<sup>a</sup> The data are average values of the two replications.

<sup>b</sup> The yield was calculated based on grams of glucose released after enzymatic hydrolysis per gram of glucose that can theoretically be produced from glucan in the substrate.

and then remained almost constant up to 72 h. The glucose yield from the untreated wood after 72 h of hydrolysis was only 5.3% of the theoretical glucose yield, while the glucose released from the pretreated wood during the first 24 h of hydrolysis ranged between 52.5% and 87.3% (Table 5). The maximum conversion yield for dilute acid and ionic liquid treated materials were 59.7% and 82.1–95.4%, respectively. Pretreatment of aspen wood with ionic liquid greatly enhanced the rate of hydrolysis and 40.2–49.9% of the cellulose in the ionic liquid treated wood was converted to glucose in just 6 h, versus 26.4% yield when the pure cellulose was hydrolyzed. Compared to the dilute acid pretreatment, the ionic liquid treatment was much more efficient in improving the yield and rate of hydrolysis. The aspen wood after 5 h ionic liquid treatment was the easiest digestible substrate with more than 95% cellulose conversion yield.

### 3.8. Effect of pretreatment on fermentability of the substrate

Fermentability of the hydrolysates was evaluated using *S. cerevisiae* which is the most important microorganism for industrial ethanol production. This strain was able to produce ethanol with theoretical ethanol yield of 87 ± 3% from glucose and a full supplemented medium. The produced ethanol from the untreated aspen wood was only 5.3% of theoretical ethanol yield after 24 h fermentation, while it was significantly higher for the pretreated materials with a 6.9–14.2-fold increase in the yield (Fig. 4). Only 42.1% of the glucan within the acid treated wood was converted to ethanol. The wood samples treated with the ionic liquid were fermented quicker than the untreated and dilute acid treated wood samples. The percentage of theoretical ethanol yield reached after

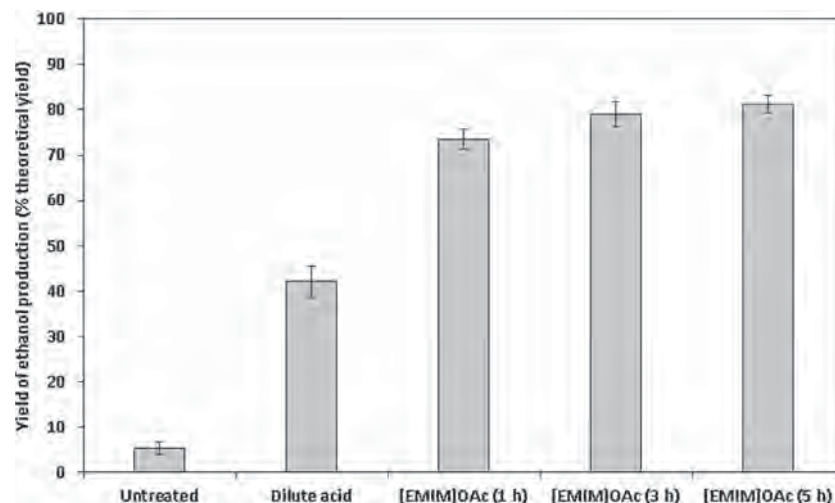
24 h fermentation, was 73.5%, 78.9%, and 81.2% for 1, 3, and 5 h ionic liquid treated wood, respectively.

### 3.9. Mass balance

A summary of the overall mass balance for the different pretreatment, enzymatic hydrolysis, and fermentation stages are presented in Table 6. Recovery of the materials after pretreatment calculated as the oven-dry mass of remained solids divided by initial oven-dried mass. The substrates were recovered at 59.5% and 76.2–81.3% after dilute acid and ionic liquid pretreatments, respectively. The maximum recovery of the materials was observed after 1 h pretreatment by the ionic liquid, whereas the highest amount of ethanol was obtained from the substrate pretreated for 5 h. Approximately, 224 g ethanol produced from 1 kg of the initial aspen wood pretreated with ionic liquid at 120 °C for 5 h, while it was only 15 and 80 g ethanol per kg of the untreated and dilute acid pretreated wood, respectively.

## 4. Discussion

Growing studies have focused on pretreatment of lignocelluloses such as wheat straw, switchgrass, corn stover, and spruce using different ionic liquids (Pooornejad, Karimi, & Behzad, 2013; Sant'Ana da Silva et al., 2011; Shafiei, Zilouei, Zamani, Taherzadeh, & Karimi, 2013; Vancov, Alston, Brown, & McIntosh, 2012). However, there is limited knowledge available regarding the impact of pretreatment on the key features of substrates that promote hydrolytic susceptibility. The present study was conducted to determine the effects of ionic liquid [EMIM][OAc] and dilute



**Fig. 4.** Yield of ethanol production (percentage of theoretical yield) after 24 h fermentation by *S. cerevisiae*.

**Table 6**Summary of the overall mass balance for different pretreatments, enzymatic hydrolysis, and fermentation stages.<sup>a</sup>

	Untreated	Pretreatment			
		Dilute acid	[EMIM]OAc (1 h)	[EMIM]OAc (3 h)	[EMIM]OAc (5 h)
Solid recovery of pretreatment	–	59.5	81.3	80.0	76.2
Produced glucose by hydrolysis <sup>b</sup>	2.8–2.9	19.5–20.8	41.9–44.2	46.2–48.2	47.2–51.0
Produced ethanol <sup>c</sup>	1.5	8.0	20.1	21.6	22.4

<sup>a</sup> All the reported values are grams per 100 g of initial aspen wood.<sup>b</sup> The ranges are the results of 24 and 48 h hydrolyses.<sup>c</sup> The data are the results of 24 h fermentation by 5 g/l *S. cerevisiae*.

acid pretreatments on physiochemical properties of aspen wood using semi-quantitative techniques. The pretreated materials were comprehensively compared using their composition, surface area, accessibility to cellulase, enzymatic digestibility, structural and morphological features.

Major improvement in enzymatic digestibility (~18-fold) of aspen wood was observed via the ionic liquid pretreatment and more than 95% of the glucose theoretical yield was achieved. Consequently, the pretreatment with [EMIM]OAc succeeded in improving the fermentability of aspen wood up to about 82% of ethanol theoretical yield after 24 h fermentation using *S. cerevisiae*. For rice straw, poplar wood, and spruce up to 88%, 92%, and 73% hydrolysis yield were reported by [EMIM]OAc pretreatment, respectively (Maki-Arvela, Anugwom, Virtanen, Sjoholm, & Mikkola, 2010; Nguyen et al., 2010; Shafiei et al., 2013).

The regenerated materials had higher level of glucan and lower levels of residual lignin compared to the untreated and dilute acid treated solids. It is acknowledged that the lignin associated with microfibrils occupy the internal space of lignocellulose and restrict the accessibility to hydrolytic enzymes. The delignification effect of ionic liquids has also been investigated for various materials such as spruce and maple woods. For example, more than 32% and 19% lignin removal from maple wood has been observed after acetate containing ionic liquid and [BMIM][MeSO<sub>4</sub>] pretreatments, respectively (Lee, Doherty, Linhardt, & Dordick, 2009). The difference in the reported values could be due to the specific interactions between solvent molecules and biomass which is directly related to the type of ions in the ionic liquid, biomass recalcitrance, and applied pretreatment conditions.

Analysis of the substrate by Simons' stain and WRV methods showed significant increase in dye adsorption capacity, adsorption ratio, and swelling ability of the ionic liquid treated materials. Direct dyes (e.g. Simons' stain dyes) have high affinities for cellulose hydroxyl groups and can be considered as molecular sensors for cellulose-containing materials. They can be assumed to dock onto the cellulose surface with their long axis parallel to and aligned with the polymer chains (Inglesby & Zeronian, 2002; Wiman et al., 2012). The two direct dyes used in the Simons' stain method are different in molecular size and binding affinity for lignocelluloses, thus distributing in the pores. The small DB (Direct Blue 1) molecules are able to penetrate into both small and large pores whereas the large DO (Direct Orange 15) molecules only diffuse into the larger pores. However, in competition, DO can easily displace the DB from the larger pores due to the stronger affinity with the hydroxyl groups (Esteghlalian et al., 2001). Noticeable increases in dye adsorption capacity (1.3–2.2-fold) and adsorption ratio (1.4–2.9-fold) after the ionic liquid pretreatment indicated major improvement in total surface area (external and internal), pore sizes, and distributions. The diameter of the DO molecules is quite similar to the diameter of the catalytic core domain of endoglucanase from *T. reesei*, and thereby a higher adsorption ratio implies that the expanded pores are large enough to accommodate the enzyme molecules. Moreover, increasing the WRVs of the samples after the ionic liquid pretreatment showed enhanced porosity (internal surface area)

of the substrate (Chandra et al., 2009). It should be noted that the addition of anti-solvent (i.e., water) to ionic solvent-biomass solution, make the system thermodynamically unstable, thus rapid phase separation occurs. Nonsolvent-induced precipitation process is believed to happen via the mechanism of spinodal decomposition which leads to finely porous structures (Lu, Yan, & Texter, 2013; Machado, Habert, & Borges, 1999; Peng et al., 2012; Xing, Peng, & Chung, 2011). It could be a possible explanation for enhanced surface area caused by ionic liquid treatment. In the case of dilute acid pretreatment, deposition and accumulation of the hydrophobic lignin on the fibers surfaces might occur and prevent water molecules from reaching the porous regions by occluding the pore structure (Donohoe, Decker, Tucker, Himmel, & Vinzant, 2008). However, more investigation is needed to support this hypothesis.

Substrate accessibility to cellulases plays an important role in enzymatic hydrolysis. Higher surface area and enlarged pores are effective parameters when they can enhance the fiber accessibility to the hydrolytic enzymes. The ionic liquid was capable of great increase in the substrate accessibility to cellulase, whereas the regenerated biomass exhibited higher binding capacity and affinity for cellulase compared to the untreated and acid treated materials. A possible explanation is dissolution of restricted cellulose within the biomass matrix and precipitation of the regenerated form on the biomass surface that provides more accessibility to the cellulose (Li et al., 2009). It should be noted that the enzyme adsorption experiments were conducted at 4 °C to prevent possible hydrolysis of the cellulosic fibers. However, adsorption process is known to be temperature dependent, possibly due to the greater binding constant at lower temperature. Therefore, the lower the temperature, the more protein adsorbs on the surface (Ooshima, Sakata, & Harano, 1983). On the other hand, at higher temperatures (e.g., 10–50 °C) hydrolysis of the fibers could impact the adsorption process by decreasing the active adsorption sites and altering the substrate features.

Recently, impact of ionic liquid pretreatment conditions on cellulose crystalline structure has been investigated for different feedstocks, including microcrystalline cellulose (Avicel) and lignocelluloses (Cheng et al., 2011; Maki-Arvela et al., 2010; Vancov et al., 2012). It has been shown that the pretreatment resulted in a loss of native cellulose crystalline structure. For Avicel, cellulose I lattice expanded and distorted prior to full dissolution in [EMIM][OAc], and upon precipitation the former structure was converted to a less ordered intermediate form, whereas fully dissolved cellulose was regenerated to a mixture of cellulose II and amorphous cellulose. However, for lignocelluloses, ionic liquid mainly disrupted the crystalline structure of cellulose and the pretreated materials exhibited typically lower degrees of crystallinity. In the current study, structural analysis of the materials was examined by FTIR spectroscopy and the ionic liquid treated materials were found to be less crystalline compared to the untreated and acid treated biomass. This could be due to the rapid precipitation during the regeneration step which prevents the dissolved cellulose recombine into the original crystalline structure (Li et al., 2009). Higher LOI value for dilute acid treated wood implied that hot dilute acid may promote some



recrystallization. It might be explained that hydrothermal conditions and partial hemicellulose removal (Table 1) could lead to annealing of cellulose and increase its crystallinity (Karimi, Shafiei, & Kumar, 2013; Li et al., 2010).

In accordance with the results of compositional analysis, the delignification effect of the ionic liquid treatment was well detected from the spectrum. Furthermore, SEM images confirmed that significant morphological modification occurred after the ionic liquid pretreatment compared to the dilute acid treatment. It is expected that the amorphous, porous, and fragmental regenerated cellulose could provide more accessible surfaces area and consequently be more amiable to the enzymatic attack.

The agreement between the digestibility and semi-quantitative analysis results implied that the applied techniques are strong indicators of the changes in substrate and enzyme–substrate factors caused by the ionic liquid and dilute acid pretreatments.

## 5. Conclusions

Assessment of the ionic liquid and dilute sulfuric acid pretreated aspen wood by rapid semi-quantitative methods showed that the treatment with [EMIM]OAc, even for a short time, resulted in significantly higher accessibility and affinity to cellulase enzyme, lower cellulose crystallinity, and improved morphological features. The aforementioned modifications caused by the ionic liquid pretreatment led to efficient digestibility of aspen wood, whereas the dilute acid was more suitable for xylan removal.

## Acknowledgements

This work was financially supported by Institute of Biotechnology and Bioengineering, Isfahan University of Technology and BioFuelNet, McGill University.

## References

- Alvira, P., Tomas-Pejo, E., Ballesteros, M., & Negro, M. J. (2010). Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. *Bioresource Technology*, *101*(13), 4851–4861.
- Auxenfans, T., Buchoux, S., Djellab, K., Avondo, C., Husson, E., & Sarazin, C. (2012). Mild pretreatment and enzymatic saccharification of cellulose with recycled ionic liquids towards one-batch process. *Carbohydrate Polymers*, *90*(2), 805–813.
- Carrillo, A., Colom, X., Sunol, J. J., & Saurina, J. (2004). Structural FTIR analysis and thermal characterisation of lyocell and viscose-type fibres. *European Polymer Journal*, *40*(9), 2229–2234.
- Chandra, R., Ewanick, S., Hsieh, C., & Saddler, J. N. (2008). The characterization of pretreated lignocellulosic substrates prior to enzymatic hydrolysis. Part 1. A modified Simons' staining technique. *Biotechnology Progress*, *24*(5), 1178–1185.
- Chandra, R. P., Bura, R., Mabee, W. E., Berlin, A., Pan, X., & Saddler, J. N. (2007). Substrate pretreatment: The key to effective enzymatic hydrolysis of lignocellulosics? *Advances in Biochemical Engineering/Biotechnology*, *108*, 67–93.
- Chandra, R. P., Esteghlalian, A. R., & Saddler, J. N. (2008). Assessing substrate accessibility to enzymatic hydrolysis by cellulases. In T. Q. Hu (Ed.), *Characterization of lignocellulosic materials* (pp. 60–80). Oxford: Blackwell Publishing Ltd.
- Chandra, R. P., Ewanick, S. M., Chung, P. A., Au-Yeung, K., Del Rio, L., Mabee, W., et al. (2009). Comparison of methods to assess the enzyme accessibility and hydrolysis of pretreated lignocellulosic substrates. *Biotechnology Letters*, *31*(8), 1217–1222.
- Cheng, G., Varanasi, P., Li, C., Liu, H., Melnichenko, Y. B., Simmons, B. A., et al. (2011). Transition of cellulose crystalline structure and surface morphology of biomass as a function of ionic liquid pretreatment and its relation to enzymatic hydrolysis. *Biomacromolecules*, *12*(4), 933–941.
- Demirbas, A. (2011). Competitive liquid biofuels from biomass. *Applied Energy*, *88*(1), 17–28.
- Diedericks, D., van Rensburg, E., Garcia-Aparicio, M. D., & Gorgens, J. F. (2012). Enhancing the enzymatic digestibility of sugarcane bagasse through the application of an ionic liquid in combination with an acid catalyst. *Biotechnology Progress*, *28*(1), 76–84.
- Donohoe, B. S., Decker, S. R., Tucker, M. P., Himmel, M. E., & Vinzant, T. B. (2008). Visualizing lignin coalescence and migration through maize cell walls following thermochemical pretreatment. *Biotechnology and Bioengineering*, *101*(5), 913–925.
- El Seoud, O. A., Koschella, A., Fidale, L. C., Dorn, S., & Heinze, T. (2007). Applications of ionic liquids in carbohydrate chemistry: A window of opportunities. *Biomacromolecules*, *8*(9), 2629–2647.
- Esteghlalian, A. R., Bilodeau, M., Mansfield, S. D., & Saddler, J. N. (2001). Do enzymatic hydrolyzability and Simons' stain reflect the changes in the accessibility of lignocellulosic substrates to cellulase enzymes? *Biotechnology Progress*, *17*(6), 1049–1054.
- Geng, X. L., & Henderson, W. A. (2012). Pretreatment of corn stover by combining ionic liquid dissolution with alkali extraction. *Biotechnology and Bioengineering*, *109*(1), 84–91.
- Ghose, T. K. (1987). Measurement of cellulase activities. *Pure and Applied Chemistry*, *59*(2), 257–268.
- Goshadrou, A., Karimi, K., & Taherzadeh, M. J. (2011). Bioethanol production from sweet sorghum bagasse by *Mucor hiemalis*. *Industrial Crops and Products*, *34*(1), 1219–1225.
- Goshadrou, A., Karimi, K., & Taherzadeh, M. J. (2013). Ethanol and biogas production from birch by NMMO pretreatment. *Biomass and Bioenergy*, *49*, 95–101.
- Gupta, R., Khasa, Y. P., & Kuhad, R. C. (2011). Evaluation of pretreatment methods in improving the enzymatic saccharification of cellulosic materials. *Carbohydrate Polymers*, *84*(3), 1103–1109.
- Inglesby, M. K., & Zeronian, S. H. (2002). Direct dyes as molecular sensors to characterize cellulose substrates. *Cellulose*, *9*(1), 19–29.
- Karimi, K., Shafiei, M., & Kumar, R. (2013). Progress in physical and chemical pretreatment of lignocellulosic biomass. In V. K. Gupta, & M. G. Tuohy (Eds.), *Biofuels technologies: Recent developments* (pp. 53–96). Berlin: Springer Science Publishers.
- Kumar, R., & Wyman, C. E. (2009). Cellulase adsorption and relationship to features of corn stover solids produced by leading pretreatments. *Biotechnology and Bioengineering*, *103*(2), 252–267.
- Lee, S. H., Doherty, T. V., Linhardt, R. J., & Dordick, J. S. (2009). Ionic liquid-mediated selective extraction of lignin from wood leading to enhanced enzymatic cellulose hydrolysis. *Biotechnology and Bioengineering*, *102*(5), 1368–1376.
- Li, C., Knierim, B., Manisseri, C., Arora, R., Scheller, H. V., Auer, M., et al. (2010). Comparison of dilute acid and ionic liquid pretreatment of switchgrass: Biomass recalcitrance, delignification and enzymatic saccharification. *Bioresource Technology*, *101*(13), 4900–4906.
- Li, Q., He, Y. C., Xian, M., Jun, G., Xu, X., Yang, J. M., et al. (2009). Improving enzymatic hydrolysis of wheat straw using ionic liquid 1-ethyl-3-methyl imidazolium diethyl phosphate pretreatment. *Bioresource Technology*, *100*(14), 3570–3575.
- Lu, J., Yan, F., & Texter, J. (2013). Advanced applications of ionic liquids in polymer science. *Progress in Polymer Science*, *34*(5), 431–448.
- Luo, X., & Zhu, J. (2011). Effects of drying-induced fiber hornification on enzymatic saccharification of lignocelluloses. *Enzyme and Microbial Technology*, *48*(1), 92–99.
- Machado, P. S. T., Habert, A. C., & Borges, C. P. (1999). Membrane formation mechanism based on precipitation kinetics and membrane morphology: Flat and hollow fiber polysulfone membranes. *Journal of Membrane Science*, *155*(2), 171–183.
- Maki-Arvela, P., Anugwom, I., Virtanen, P., Sjöholm, R., & Mikkola, J. P. (2010). Dissolution of lignocellulosic materials and its constituents using ionic liquids – A review. *Industrial Crops and Products*, *32*(3), 175–201.
- Moniruzzaman, M., & Ono, T. (2012). Ionic liquid assisted enzymatic delignification of wood biomass: A new 'green' and efficient approach for isolating of cellulose fibers. *Biochemical Engineering Journal*, *60*, 156–160.
- Mora-Pale, M., Meli, L., Doherty, T. V., Linhardt, R. J., & Dordick, J. S. (2011). Room temperature ionic liquids as emerging solvents for the pretreatment of lignocellulosic biomass. *Biotechnology and Bioengineering*, *108*(6), 1229–1245.
- Nelson, M. L., & O'Connor, R. T. (1964). Relation of certain infrared bands to cellulose crystallinity and crystal lattice type. Part II. A new infrared ratio for estimation of crystallinity in celluloses I and II. *Journal of Applied Polymer Science*, *8*(3), 1325–1341.
- Nguyen, T. A., Kim, K. R., Han, S. J., Cho, H. Y., Kim, J. W., Park, S. M., et al. (2010). Pretreatment of rice straw with ammonia and ionic liquid for lignocellulose conversion to fermentable sugars. *Bioresource Technology*, *101*(19), 7432–7438.
- Nigam, P. S., & Singh, A. (2011). Production of liquid biofuels from renewable resources. *Progress in Energy and Combustion Science*, *37*(1), 52–68.
- Oh, S. Y., Yoo, D. I., Shin, Y., Kim, H. C., Kim, H. Y., Chung, Y. S., et al. (2005). Crystalline structure analysis of cellulose treated with sodium hydroxide and carbon dioxide by means of X-ray diffraction and FTIR spectroscopy. *Carbohydrate Research*, *340*(15), 2376–2391.
- Ooshima, H., Sakata, M., & Harano, Y. (1983). Adsorption of cellulase from *Trichoderma viride* on cellulose. *Biotechnology and Bioengineering*, *25*(12), 3103–3114.
- Peng, N., Widjojo, N., Sukitpaneinit, P., Teoh, M. M., Lipscomb, G. G., Chung, T., et al. (2012). Evolution of polymeric hollow fibers as sustainable technologies: Past, present, and future. *Progress in Polymer Science*, *37*(10), 1401–1424.
- Pinto, J. H., & Kamden, D. P. (1996). Comparison of pretreatment methods on the enzymatic saccharification of aspen wood. *Applied Biochemistry and Biotechnology*, *61*(3), 289–297.
- Poornejad, N., Karimi, K., & Behzad, T. (2013). Improvement of saccharification and ethanol production from rice straw by NMMO and [BMIM][OAc] pretreatments. *Industrial Crops and Products*, *41*, 408–413.
- Program, N. S. (2000). *Water retention value of chemical pulps*. Stockholm, Sweden: Scandinavian Pulp, Paper and Board Testing Committee. SCAN-C 62:00.
- Sanchez, Ó. J., & Cardona, C. A. (2008). Trends in biotechnological production of fuel ethanol from different feedstocks. *Bioresource Technology*, *99*(13), 5270–5295.
- Sant'Ana da Silva, A., Lee, S. H., Endo, T., & Bon, E. P. (2011). Major improvement in the rate and yield of enzymatic saccharification of sugarcane bagasse via pretreatment with the ionic liquid 1-ethyl-3-methylimidazolium acetate ([EMIM][Ac]). *Bioresource Technology*, *102*(22), 10505–10509.



- Seeta Laxman, R., & Lachke, A. H. (2009). Bioethanol from lignocellulosic biomass. Part I. Pretreatment of the substrates. In A. Pandey (Ed.), *Handbook of plant-based biofuels* (pp. 121–140). Boca Raton: CRC Press.
- Shafiei, M., Zilouei, H., Zamani, A., Taherzadeh, M. J., & Karimi, K. (2013). Enhancement of ethanol production from spruce wood chips by ionic liquid pretreatment. *Applied Energy*, *102*, 163–169.
- Sluiter, A., Hames, B., Ruiz, R., Scarlata, C., Sluiter, J., Templeton, D., et al. (2008). *Determination of structural carbohydrates and lignin in biomass*. National Renewable Energy Laboratory, NREL/TP-510-42618.
- Taherzadeh, M. J., & Karimi, K. (2007). Enzyme-based hydrolysis processes for ethanol from lignocellulosic materials: A review. *Bioresources*, *2*(4), 707–738.
- Taherzadeh, M. J., & Karimi, K. (2008a). Bioethanol: Market and production processes. In A. Nag (Ed.), *Biofuels refining and performance* (pp. 69–106). New York: McGraw-Hill.
- Taherzadeh, M. J., & Karimi, K. (2008b). Pretreatment of lignocellulosic wastes to improve ethanol and biogas production: A review. *International Journal of Molecular Sciences*, *9*(9), 1621–1651.
- Tan, H. T., & Lee, K. T. (2012). Understanding the impact of ionic liquid pretreatment on biomass and enzymatic hydrolysis. *Chemical Engineering Journal*, *183*, 448–458.
- Uju, Shoda, Y., Nakamoto, A., Goto, M., Tokuhara, W., Noritake, Y., et al. (2012). Short time ionic liquids pretreatment on lignocellulosic biomass to enhance enzymatic saccharification. *Bioresource Technology*, *103*(1), 446–452.
- Vancov, T., Alston, A. S., Brown, T., & McIntosh, S. (2012). Use of ionic liquids in converting lignocellulosic material to biofuels. *Renewable Energy*, *45*, 1–6.
- Wiman, M., Dienes, D., Hansen, M. A. T., van der Meulen, T., Zacchi, G., & Lidén, G. (2012). Cellulose accessibility determines the rate of enzymatic hydrolysis of steam-pretreated spruce. *Bioresource Technology*, *126*, 208–215.
- Xing, D. Y., Peng, N., & Chung, T. (2011). Investigation of unique interactions between cellulose acetate and ionic liquid [EMIM]SCN, and their influences on hollow fiber ultrafiltration membranes. *Journal of Membrane Science*, *380*(15), 87–97.
- Yang, B., & Wyman, C. E. (2008). Pretreatment: The key to unlocking low-cost cellulosic ethanol. *Biofuels, Bioproducts and Biorefining*, *2*(1), 26–40.
- Zhao, H., Jones, C. I. L., Baker, G. A., Xia, S., Olubajo, O., & Person, V. N. (2009). Regenerating cellulose from ionic liquids for an accelerated enzymatic hydrolysis. *Journal of Biotechnology*, *139*(1), 47–54.
- Zhu, S., Wu, Y., Chen, Q., Yu, Z., Wang, C., Jin, S., et al. (2006). Dissolution of cellulose with ionic liquids and its application: A mini-review. *Green Chemistry*, *8*(4), 325–327.

# Technology and Product Reports

## Design, Construction, and Operation of a Demonstration Rainwater Harvesting System for Greenhouse Irrigation at McGill University, Canada

Sadman Islam<sup>1</sup>, Mark Lefsrud<sup>2,3</sup>, Jan Adamowski<sup>2</sup>,  
Blake Bissonnette<sup>1</sup>, and Allison Busgang<sup>1</sup>

**ADDITIONAL INDEX WORDS.** water recycling, water supply, water quality, urban water demand, sustainability, algae growth

**SUMMARY.** Increasing stress on urban water demand has led to the exploration of the potential of rainwater use and water recycling to promote sustainable water resources management. Rainwater harvesting (RWH) not only has the potential to reduce water demand but also contributes to other sustainable objectives, including reducing stormwater pollutant loads, reducing erosion, and inducing natural flow regimes by means of flood control, in urban streams. This research involved the design, construction, and field-testing of an RWH system used to irrigate greenhouses at the Macdonald Campus of McGill University in Quebec, Canada. The purpose of the RWH system was to collect rainwater from a roof area of  $\approx 610 \text{ m}^2$  (the Horticulture Services Building on the Macdonald Campus of McGill University) to meet the irrigation demands of the two Horticulture Research Center greenhouses on the campus ( $\approx 149 \text{ m}^2$  each) from May to October. Over its two years of operation, it was found that the amount of rainwater collected did not only meet the peak irrigation demands of the greenhouses (which amounted to almost 700 gal of water per day), but that there was also enough water for the irrigation of the nearby student-run gardens. The harvested rainwater was clear and did not cause any harm to the plants. The major problem that was experienced during the operation of the RWH system was that of algae growth in one of the water collection tanks. This issue was resolved by covering the tank with metallic green wallpaper, thereby blocking most of the sunlight from entering the tank. The RWH system is currently being used for irrigation and as a demonstration project to promote the learning of sustainable technologies on campus and in the surrounding communities.

We thank the McGill Sustainability Projects Fund for providing us with the required funding for the setup and maintenance of the project.

Department of Bioresource Engineering, McGill University, 21111 Lakeshore Road, Sainte-Anne-de-Bellevue, H9X 3V9, QC, Canada

<sup>1</sup>Research Assistant.

<sup>2</sup>Assistant Professor.

<sup>3</sup>Corresponding author. E-mail: mark.lefsrud@mcgill.ca.

As a part of the family of green technologies, RWH can be defined as the capturing of rain runoff from roofs and other surfaces and storing it for a later purpose (Despins et al., 2009). As an ancient practice, RWH cisterns were common in ancient Greek, Etruscan, Roman, Indian, and other civilizations (Boers

and Asher, 1982). In Jordan for example, surface runoff has been collected for over 4000 years. Elsewhere, archaeological research in Venice resulted in the identification of more than 6000 subterranean rainwater cisterns constructed during the Middle Ages for domestic water supply. More recently, the advent of urban sprawl has resulted in a decrease in the amount of forested lands, wetlands, and other forms of open spaces that absorb and clean storm water in the natural system (Leopold, 1968). This has caused degradation in the water quality of water bodies that are now used in the agricultural and domestic sectors. The practice of RWH has been gaining popularity as the usage of rainwater is much cleaner (in terms of carbon dioxide emissions) than the usage of municipal water supplies. Like other water conservation techniques, RWH is considered to be a viable means to manage urban water resources more efficiently and sustainably (Basinger et al., 2010).

Domestic water usage is a significant component of the global water demand. RWH can be used for both nonpotable and potable purposes such as garden use, toilet flushing, washing clothes, hot water systems, and drinking water supply (Khastagir and Jayasuriya, 2010). Catchment area, storage material, and the distribution system are a few design considerations that have to be taken into account when constructing a RWH system. When selecting a rainwater tank, a house owner often only focuses on the location where the tank will be placed, its aesthetics, and the cost. However, other key design variables that need to be considered are: amount of precipitation in the area, extent of catchment area, and the end use of the water. If the tank is sized properly, the volume of rainwater in the tank will be able to supplement the household water demand; this also reduces the chances of the tank being empty or overflowing. Also, if the catchment area is fairly large, a greater number of end-use applications can allow for greater water savings.

According to the United Nations Environment Programme (UNEP, 2012), examples of RWH and its utilization can be found all across the world. For example, with almost 86% of Singapore's population living in high-rise buildings, using RWH has

become an important component of reducing rising urban water demand in the region. In cities and regions such as Tokyo, Thailand, and China, RWH is seen as a way of mitigating water shortages, controlling floods, and securing water for emergencies. In Bangladesh, where the rural population is plagued by arsenic in the drinking water, RWH is being used to provide a potable water supply, and in St. Thomas, U.S. Virgin Islands it is now mandatory to install a RWH system to acquire a residential building permit.

In Canada, a study in Ontario by Farahbakhsh et al. (2009) indicated that if there are no weather anomalies and the precipitation pattern follows the historical trend, then municipal water demand can be reduced by as much as 47% in domestic households in Ontario, Canada, if RWH is used for domestic water supply. In southeast Brazil, Ghisi (2006) estimated that RWH practices can reduce potable water demand by 48% to 100%. In Germany, where RWH has been in use since the 1980s, potable water demand has decreased by 30% to 60% due to the use of roof runoff harvested in 4- to 6-m<sup>3</sup> tanks, which is then used for toilet flushing (Hermann and Schmida, 1999). RWH is being used in the rural areas of the developing world (such as India, China, and Africa), where it is mainly used for irrigation during dry periods. In such environments, RWH is seen as a measure that can help fight food scarcity (Fox et al., 2005).

The quality of water collected in a RWH system is affected by several factors, which include the proximity to roads and heavy industries, presence of wildlife, and meteorological conditions such as temperature and rainfall patterns in the region (Despins et al., 2009). Although these factors may deteriorate the quality of the collected rainwater, Kumar (2004) states that treatment by prestorage treatment devices such as filtration or first flush diversion can help to expel most of the harmful sediments from the roof (catchment area) into the surroundings and also prevent the stagnation of water in the tanks, thereby restricting the breeding of insects. Similarly, poststorage treatment devices such as ultraviolet disinfection, chlorination, or slow sand filtration can further improve the water quality,

in some cases making the harvested rainwater a source of potable water. As an added precaution, storage tanks must be closed at all times to prevent the entry of insects and to reduce evaporation losses. Finally, the most commonly occurring problem of algal blooms in storage tanks can be mitigated by both chemical and physical techniques (Bartsch, 1954). Chemical techniques include the use of copper sulfate in small concentrations such that the algal bloom is destroyed without harming the roots of the plants (if the water is used for irrigation). Even though the solution is effective, it is not viable as the cost of implementation rises with increasing volumes of water in the storage tank. Covering the storage tanks with a reflective material, which prevents sunlight transmission into the water tanks, is an alternative solution (Bartsch, 1954). This is a significantly less expensive and more practical approach. Another issue that must be addressed in the design of RWH systems is the handling of overflows during large rainfall events. Methods for handling overflow may include onsite infiltration, as well as discharging to an existing storm sewer infrastructure (Farahbakhsh et al., 2009).

The cost of a RWH system varies depending on the size of the catchment area, the type of cisterns chosen to store the rainwater, and the piping material used to transport the water for its end use. Cost was by far the biggest barrier identified for the implementation of RWH systems in a recent study conducted by Farahbakhsh et al. (2009); the typical roof area of a domestic household in Canada is  $\approx 230$  m<sup>2</sup> (Canadian Mortgage and Housing Corp., 2012), which translates to a cost in the range of C\$2000

to C\$8000 (in Canadian dollars) or the setup of a RWH system. The yearly maintenance costs and decommissioning cost at the end of the lifespan of the systems costs an additional C\$2000 (Roebuck and Ashley, 2006). It should be noted that RWH systems have one of the lowest payback periods for sustainable engineering systems (maximum 15 years), and government incentives can often offset some of the expenses (Despins et al., 2009).

The Faculty of Agricultural and Environmental Sciences on the Macdonald Campus of McGill University has been active in the promotion of sustainable technologies. The Horticulture Research Center at the Macdonald Campus serves as a training, education, and research center for students involved in vegetable and fruit production and is used by numerous departments on the campus. The Horticulture Services Building is used as a storage, processing, and retail space for the activities that occur in the surrounding gardens and the two greenhouses. Irrigation of the greenhouses consumes  $\approx 700$  gal of freshwater per day. The objective of the RWH project described in this article is to manage urban water resources more efficiently and promote the learning of practical sustainable technologies at the university and the surrounding communities through educational tours to the project site. Although research on domestic RWH systems has been increasing, there are still relatively few publications that describe the construction and advantages of such systems. This article is intended to fill this gap by explaining cost-effective ways for the setup and maintenance of this sustainable water supply option.

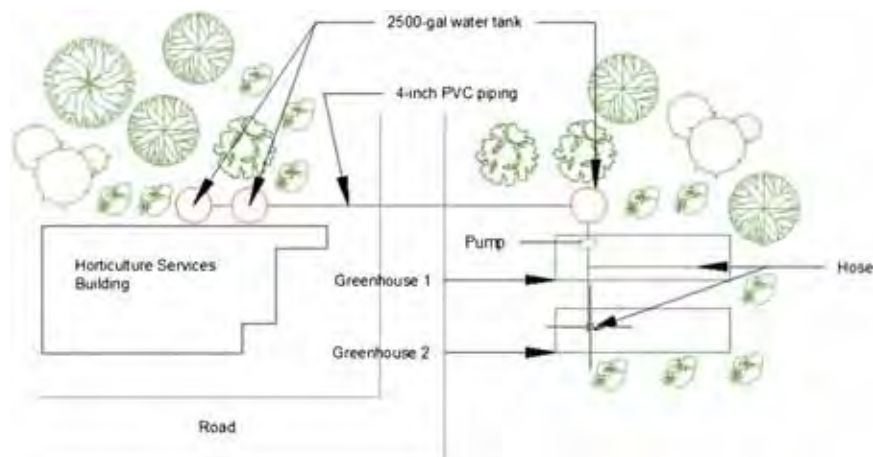
### Units

To convert U.S. to SI, multiply by	U.S. unit	SI unit	To convert SI to U.S., multiply by
29.5735	fl oz	mL	0.0338
0.3048	ft	m	3.2808
0.0929	ft <sup>2</sup>	m <sup>2</sup>	10.7639
0.0283	ft <sup>3</sup>	m <sup>3</sup>	35.3147
3.7854	gal	L	0.2642
0.7457	horsepower	kW	1.3410
2.54	inch(es)	cm	0.3937
25.4	inch(es)	mm	0.0394
1	ppb	$\mu\text{g}\cdot\text{L}^{-1}$	1
1	ppm	$\text{mg}\cdot\text{L}^{-1}$	1
6.8948	psi	kPa	0.1450
$(^{\circ}\text{F} - 32) \div 1.8$	$^{\circ}\text{F}$	$^{\circ}\text{C}$	$(^{\circ}\text{C} \times 1.8) + 32$

## Methodology

The setup of the RWH system (Fig. 1) comprises the Horticulture Services Building and the two Horticulture Research Center greenhouses ( $\approx 50.3$  m away). Rainwater is collected from a sheet metal roof with a catchment area of  $\approx 610$  m<sup>2</sup>. Roof runoff is redirected using 4-inch aluminum eaves troughs into a 2500-gal polyethylene water tank (part no. 40051; Norwesco, St. Bonifacius, MN), with the help of downspouts. An aluminum mesh attached to the top of the eaves troughs prevents leaves and other large debris from entering the water tank. Weather data collected from Environment Canada helped in the sizing of the water tanks by determining the average annual precipitation pattern along with the longest recorded dry period on the island of Montreal. The average annual amount of rainfall in Montreal over the period of 1971–2000 is 834.9 mm, whereas the most prolonged interval without precipitation is 7 d over the past five years (Environment Canada 2012a, 2012b). However, the system was designed to withstand a 1 in 10 year storm event, resulting in the accumulation of 19 m<sup>3</sup> of rainwater over a period of 15 min. It is estimated that the tanks require 52 mm rainfall to fill up with a collective storage volume of 7500 gal, able to sustain the peak irrigation demands of both greenhouses for 11 d.

The collection tanks (Fig. 2) are located next to the Horticulture Services Building. The site is covered with vegetation, and the soil contains a large amount of rocks, adding both aesthetics and stability to the system. The first tank is connected in series to two other identical tanks, each at a decreasing elevation. Gravitational pull and the accumulated head pressure in each tank govern the flow of water between the tanks. The tanks are connected to each other via 4-inch polyvinyl chloride (PVC) pipes, with a system of 4-inch valves and overflows allowing one to control the inflow and outflow of water in each tank. The third tank rests beside the greenhouses (which have an area of  $\approx 149$  m<sup>2</sup> each), and it is connected to a 1-horsepower centrifugal water pump (part no. 4011K11; McMaster-Carr, Santa Fe Springs, CA) that supplies irrigation water to both greenhouses.



**Fig. 1.** Overview map of the rainwater harvesting (RWH) system at the Horticulture Services Building (Macdonald Campus, Sainte-Anne-de-Bellevue, QC, Canada) and at the two Horticulture Research Center greenhouses. Rainwater is collected from a sheet metal roof, directed using 4-inch (10.2 cm) aluminum eaves troughs into a three 2500-gal (9463.5 L) polyethylene water tank.



**Fig. 2.** A side view of the Horticulture Services Building (Macdonald Campus, Sainte-Anne-de-Bellevue, QC, Canada), two 2500-gal (9463.5 L) polyethylene water tanks, and the sheet metal roof directed using 4-inch (10.2 cm) aluminum eaves troughs with downspouts.



**Fig. 3.** A view of the pressure switch at the Horticulture Services Building (Macdonald Campus, Sainte-Anne-de-Bellevue, QC, Canada) that turns the 1-horsepower (0.75 kW) centrifugal water pump on and off for irrigation, when the line pressures are 25 and 40 psi (172.4 and 275.8 kPa), respectively.



With an inlet diameter of 2 inches and an outlet of 1.5 inches, the 1-horsepower centrifugal water pump delivers water at a flow rate of 22 gal/min and 40 psi of pressure at the furthest end of the greenhouses (24 m away). The pump is completely automated with the help of a pressure switch (Fig. 3) that turns the pump on and off when the line pressures are 25 and 40 psi, respectively. The pump helps to meet the peak irrigation needs of the greenhouses from May to October, which can amount up to  $\approx 350$  gal of water per day in each greenhouse. The piping within the greenhouses mainly consists of 1.5-inch semirigid pipes, which allows for flexibility within the structures.

Overflows are designed on all three tanks with 4-inch PVC pipes. The overflows help to discharge the water into the surroundings once the water volume in the tanks reaches 2200 gal. The overflows in the first two tanks are combined together with a “T-junction,” and the water is discharged into the nearby vegetation. The entire system performs efficiently, since it requires no manual labor from the beginning (collection of rainwater) to the end (irrigation of plants).

To verify the chemical and biological composition of the collected rainwater, a thorough water analysis was carried out. Two rainwater samples were collected in 25-mL containers from the first and second storage tanks. The samples were collected 1 week after a rainstorm event (25 Sept. 2012) and stored in ambient temperature ( $\approx 25$  °C). A third sample was collected from the tap that was previously used to supply irrigation water to the two greenhouses, for comparison. The samples were bottled and shipped to a water analysis laboratory that analyzed the samples for various heavy metals as well as the total Kjeldahl nitrogen content using inductively coupled plasma mass spectrometry. Three water samples were collected and stored in the same fashion for the bacterial count. The samples were shipped to the same water analysis laboratory where the total bacterial count was carried out using different plating techniques.

## Results and discussion

This project was the first of its kind to be undertaken at McGill University and serves as a tool for

both irrigation and education. Similar projects can be found at the University of Guelph and McMaster University in Canada, where research is being done to design and operate household RWH systems for flushing toilets, watering gardens, and drinking purposes. To date, the RWH system at the Macdonald Campus of McGill University has been successful throughout the summer and fall in meeting the irrigation demands of both greenhouses combined. Before the RWH system, the town of Sainte-Anne-de-Bellevue, QC, Canada, mainly supplied the irrigation water with some water provided by the Macdonald Campus water supply system (<10%). A decentralized water supply that uses its own source of water is a more sustainable way of providing water for irrigation purposes at the Macdonald Campus. In addition, rainwater is naturally soft water without any minerals, chlorine, fluoride, and other chemicals.

The RWH system not only stores enough rainwater for watering the greenhouses but it has also been able to water the nearby student-run gardens. The design includes steps for collection, filtering, storage, and delivery of the water for irrigation. Although the system did not include steps for the chemical treatment of the stored rainwater, the water that was used for irrigation in the greenhouses was clear and did not have any odor associated with it. Since the

catchment area was a sheet metal roof, it limited the leaching of harmful substances such as lead, mercury, and polycyclic aromatic hydrocarbons into the captured rainwater that would have otherwise been present if the roof was made of asphalt shingles (Van Metre and Mahler, 2003). At the same time, the presence of dissolved and particulate copper is minimized due to the absence of clay or concrete tiles on the roof (Forster, 1996).

Storage tank material can also affect the pH of rainwater that has been collected and change its chemical quality (Despins et al., 2009). The use of high-density polyethylene water tanks in the setup was because plastic tanks create no changes in the pH and chemical composition of the stored water (Hart and White, 2006). Concrete cisterns have reported a rise in the pH from 5 on the roof surface to 9.4 in the tank (Hart and White, 2006). There is also a significant concern about chemical leaching of zinc from metal tanks. In the RWH system at McGill University, sedimentation has played a major role in reducing the contaminant load of the stored rainwater. Sedimentation allows the water to rest in order for the flocculated or coagulated particles to settle out. Ideally the water should be allowed to rest for more than 4 h for the sedimentation process to be effective. The collected rainwater was allowed to rest for 24 h after every

**Table 1. Comparison of levels of heavy metals in rainwater (collected on 25 Sept. 2012, from the roof of the Horticulture Services Building, Macdonald Campus, Sainte-Anne-de-Bellevue, QC, Canada) and a municipal tap water sample (at the same location) to determine water quality of the rainwater harvested for greenhouse irrigation purposes. Samples were collected 1 week after a rainstorm event, stored at ambient temperature [ $\approx 25$  °C (77.0 °F)], and analyzed for various heavy metals using inductively coupled plasma mass spectrometry.**

Parameter	Unit <sup>z</sup>	Reported detection			
		limit	Rain 1 <sup>y</sup>	Rain 2 <sup>y</sup>	Tap 1 <sup>x</sup>
Copper	$\mu\text{g}\cdot\text{L}^{-1}$	3.0	<3.0	<3.0	9.3
Lead	$\mu\text{g}\cdot\text{L}^{-1}$	1.0	<1.0	<1.0	<1.0
Zinc	$\mu\text{g}\cdot\text{L}^{-1}$	3.0	285	359	7.4
Aluminum	$\mu\text{g}\cdot\text{L}^{-1}$	30	<30	<30	45
Iron	$\mu\text{g}\cdot\text{L}^{-1}$	300	<300	<300	<300
Manganese	$\mu\text{g}\cdot\text{L}^{-1}$	5	<5	<5	<5
Sodium	$\mu\text{g}\cdot\text{L}^{-1}$	2000	<2000	<2000	9720
Selenium	$\mu\text{g}\cdot\text{L}^{-1}$	2	<2	<2	<2
Total mercury	$\text{mg}\cdot\text{L}^{-1}$	0.0001	<0.0001	<0.0001	<0.0001

<sup>z</sup>1  $\mu\text{g}\cdot\text{L}^{-1}$  = 1 ppb, 1  $\text{mg}\cdot\text{L}^{-1}$  = 1 ppm.

<sup>y</sup>Two rainwater samples (Rain1 and Rain2) were collected in 25-mL (0.85 fl oz) containers from the first and second storage tanks.

<sup>x</sup>Sample (Tap 1) was collected from the tap that was previously used to supply irrigation water to the two greenhouses, for comparison. The sample was bottled in a 25-mL container.



rainstorm before it was used for irrigation. This provided a source of primary treatment to the collected water.

Water quality also is significantly affected by the type of piping material that is used to transport the water for its end use. PVC pipes were used in this project since they are relatively inert and do not affect the water quality in any significant way. In a comparison between concrete and PVC pipes, Davies et al. (2010) found that concrete drainage systems have a significant influence on water chemistry, particularly where the inflow is acidic. The major factor identified in the research was the presence of calcium, bicarbonate, and potassium ions from the concrete pipes.

As can be seen in Table 1, the chemical composition of the harvested rainwater is fairly similar to tap water. Among the various heavy metal concentrations analyzed, a notable difference can be found between the concentrations of zinc and copper (which are micronutrients for plants). Since the catchment area is made of sheet metal, the average concentration of zinc in the two rainwater samples is  $322 \mu\text{g}\cdot\text{L}^{-1}$ . Although this is  $\approx 43.5$  times higher than the zinc concentration in the tap water sample, the Canadian Council of Ministers of the Environment (CCME, 2011) states that the threshold limit for zinc in irrigation water is  $1000 \mu\text{g}\cdot\text{L}^{-1}$  when the soil pH is below 6.5. Thus, the zinc concentration in the rainwater is not toxic to the greenhouse plants. Copper concentrations however are on average three times lower in the rainwater when compared with tap water. This is can be mainly attributed to the fact that copper pipes are used as a means of delivering the municipal water to the greenhouses, whereas the stored rainwater is transferred to the greenhouses with the help of PVC and plastic piping. Plant deficiency in copper can easily be avoided by fertilization.

Table 2 displays the total nitrogen content and bacterial count for the rainwater and tap water samples. The total nitrogen content in both samples was the same, but the bacterial count in the rainwater was  $\approx 473$  times higher, when compared with the tap water. This indicates that the rainwater cannot be used as a potable water supply, but it does not make

a difference when it is solely used for irrigation.

The tanks in this study were sized to incorporate a suitable safety factor to prevent perpetual overflows and the flooding of the project site. It was calculated that a combined storage volume of 7500 gal would be able to withstand a 1 in 10 year rainstorm event lasting 15 min. Sizing the tanks to have a large storage volume also ensured that the Horticulture Research Center will not deplete its supply of irrigation water during the summer months. However, even with such a considerable storage capacity, some of the water in the tanks had to be expelled (through the overflows) since the tanks did not have enough volume to store the large amounts of water that were sometimes obtained (after big rainstorms). Since the irrigation water demand in the greenhouses declined from September onwards, unused rainwater in the storage tanks accounted for the overflowing of the cisterns. The approximation of the tanks being able to survive a 1 in 10 year rainstorm for 15 min was therefore an underestimate. Addition of another 2500-gal tank (amounting to a total storage volume

to 10,000 gal) should solve the problem of overflows.

There were no problems with the breeding of insects and the entry of foreign objects into the water tanks, as they were sealed with the plastic caps that were provided by the manufacturers. Overall, the cost of the project was  $\approx \text{C}\$8000$  (Table 3). Most of this cost was associated with the installation of the eaves troughs on the roof of the Horticultural Services Building. For ordinary domestic households, this cost would be significantly lower, especially if the eaves troughs are already installed on the roof. For an average household with a roof area of  $230 \text{ m}^2$ , the cost of installation may range between  $\text{C}\$1500$  and  $\text{C}\$2000$  (Save the Rain Campaign, 2010). This cost may change depending on the historic precipitation data of the region and the end use of the harvested rainwater. If new eaves troughs need to be installed, the cost could increase to  $\approx \text{C}\$3000$  to  $\text{C}\$5000$ .

Based on the 2011 annual precipitation data collected from a local weather station,  $411.5 \text{ m}^3$  of rainwater (including overflows) was collected by the RWH system from May to

**Table 2. Comparison of total Kjeldahl nitrogen (TKN) content and the bacterial count (BHAA) in the harvested rainwater (collected on 25 Sept. 2012, from the roof of the Horticulture Services Building, Macdonald Campus, Sainte-Anne-de-Bellevue, QC, Canada) and a municipal tap water sample (at the same location) to determine water quality of the rainwater harvested for the greenhouse irrigation purposes. Samples were collected 1 week after a rainstorm event, stored at ambient temperature [ $\approx 25^\circ\text{C}$  ( $77.0^\circ\text{F}$ )], and analyzed for TKN and BHAA (using dilution plating techniques).**

Parameter	Unit <sup>a</sup>	Rain 1 <sup>b</sup>	Rain 2 <sup>b</sup>	Tap 1 <sup>b</sup>
TKN	$\text{mg}\cdot\text{L}^{-1}$	<1.0	<1.0	<1.0
BHAA	cfu/mL	120,000	230,000	370

<sup>a</sup>1  $\text{mg}\cdot\text{L}^{-1}$  = 1 ppm, 1 cfu/mL = 29.5735 cfu/fl oz.

<sup>b</sup>Two rainwater samples (Rain 1 and Rain 2) and one tap water sample (Tap 1) were collected in 25-mL (0.85 fl oz) containers from the first and second storage tanks along with the tap that previously supplied irrigation water to the two greenhouses.

**Table 3. A detailed cost analysis for the setup of the rainwater harvesting system at the Horticulture Services building located on the Macdonald Campus of McGill University, Sainte-Anne-de-Bellevue, QC, Canada.**

Type of cost	Value (Canadian dollars)
Purchase and installation of eaves troughs	2000
Purchase of 200 ft (61.0 m) of 4-inch (10.2 cm) PVC (PVC) pipes	1000
Purchase of three 2500-gal (9463.5 L) water tanks	3300
Purchase of water pump	300
Purchase of pressure switch	36
Purchase of 100 ft (30.5 m) of semirigid piping	150
Pump and pipe fittings	1200

October. Since the irrigation of the greenhouses required  $\approx 318 \text{ m}^3$  of water during this period of time, all of the irrigation water was supplied by the RWH system. In the city of Montreal, water is priced at a very low rate of C\$0.22/ $\text{m}^3$ . If all of the collected water was used for irrigation, it would lead to a saving of C\$90.53 every year, which results in a payback period of  $\approx 88$  years. Although this is a very long period of time, for water-scarce regions where the price of water is much higher (such as Las Vegas, NV, or the Republic of Cyprus) the payback period will be much faster.

The major problem that was faced in the project was that of algae growth in the final storage tank. The first two tanks were located behind the Horticulture Services Building and were not subjected to large amounts of sunlight throughout the day, reducing problems with algae growth.

However, the final tank was situated in an open space beside the greenhouses, maximizing its exposure to sunlight, particularly during dawn and midday. Although the high-density polyethylene wall of the tank was 1-inch thick, almost 20% of the sunlight was still transmitted to the stored rainwater (MS-01; Apogee Instruments,

North Logan, UT). This created an algal bloom (that was restricted to the surface of the stored water). Before the algae could spread throughout the water body, the tank was flushed and refilled. The tank was then completely wrapped with metallic green wallpaper (as shown in Fig. 4), which helped to reflect almost all the sunlight, preventing any algal blooms. The tank has to be re-wrapped every year due to wear and tear of the wallpaper.

Fresh water scarcity is a major concern around the world, with many regions beginning to lack the water resources necessary to meet the demands of ever growing populations, as well as ecosystems. This project was able to introduce and demonstrate a simple method that can help to address the problem of freshwater scarcity. RWH systems require no treatment of the water (if it is used for irrigation), thereby reducing installation and operational costs. For this project, municipal water for greenhouse irrigation is no longer purchased from the town of Sainte-Anne-de-Bellevue, Quebec, thereby creating a savings of  $\approx$ C\$70 from the month of May to October.

Finally, an additional important outcome was the implementation of a simple and sustainable technology to increase awareness of

RWH on the McGill campus and its surroundings. The project is mainly being used as a demonstration tool to enrich learning on and around the campus on the topic of simple sustainable water technologies. The project helps to show that the setup of a RWH system is relatively simple and is an example of how individual stakeholders can begin to help the transition to more sustainable and decentralized water supply systems. The project allows visitors to appreciate that along with economic incentives (especially in regions where water prices are very high), RWH systems also provide other advantages such as reducing flooding and erosion, as well as decreasing the amount of water that needs to be produced via water treatment plants. Since the completion of the setup, a number of classes from McGill University (as well as outside visitors) have visited the site, including the following courses: BREE-327 (Bioenvironmental Engineering), BREE-510 (Watershed Systems Management), and PLNT-312 (Urban Horticulture).

## Conclusion

Various measures to meet urban water demand and to help communities transition to sustainable water use are currently being explored around the world. For example, rainwater and reclaimed wastewater have been used for a variety of purposes ranging from potable water supply to irrigation and greywater recycling used in toilets and for household cleaning. Rainwater harvesting on a small scale provides an example of one approach to help communities transition to more sustainable water resources management. The effectiveness of RWH systems mainly depend on their geographic locations due to varying amounts of precipitation in different places, the size of the catchment area, and the end use of the harvested rainwater. The Macdonald Campus of McGill University functioned as an ideal study site for such a project. Although the setup of a RWH system in Montreal has very little economic incentives, this may not be the case in water-scarce regions. The system at the Horticulture Research Center was designed to withstand a 1 in 10 year storm event lasting 15 min, to prevent any water shortage that could be



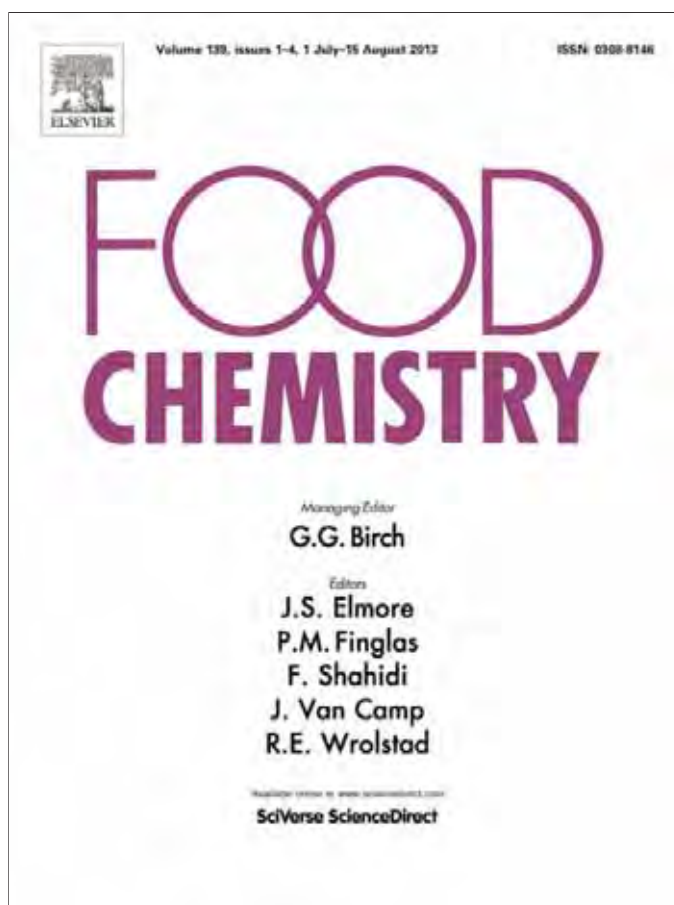
Fig. 4. A 2500-gal (9463.5 L) polyethylene water tank at the Horticulture Services Building (Macdonald Campus, Sainte-Anne-de-Bellevue, QC, Canada) wrapped with metallic green wallpaper to reflect almost all the sunlight, preventing any algal blooms.

faced during irrigation. With a catchment area of 610 m<sup>2</sup>, the RWH system located beside the Horticulture Research Center was able to supply all of the water required for irrigation of the two 149-m<sup>2</sup> greenhouses on the Macdonald Campus of McGill University from the beginning of the summer to the end of the fall. Due to the materials used for the roof (the catchment area), the storage tanks, and the piping, the collected rainwater was clean with very little change in the water chemistry. This made the harvested rainwater ideal for irrigation purposes. The only obstacle that was faced during the operation of the RWH system was algae growth in the final water tank. This issue was resolved by flushing the tank and wrapping it with metallic green wallpaper that reflected most of the sunlight away from the stored water. Not only is the system beneficial for the Horticulture Research Center and the student-run gardens, it also plays an important part in educating the student population at McGill University and the surrounding communities about the benefits of sustainable water resources technologies. Moreover, the RWH system can be expanded in the future to include other buildings located on campus to supply the buildings with their own source of greywater for toilet flushing, floor cleaning, and other purposes. As urbanization continues around the world, RWH provides a useful approach to save water resources.

### Literature cited

- Bartsch, A.F. 1954. Practical methods for control of algae and water weeds. *Public Health Rpt.* 68:749–757.
- Basinger, M., F. Montalto, and U. Lall. 2010. A rainwater harvesting system reliability model based on nonparametric stochastic rainfall generator. *J. Hydrol. (Amst.)* 392:105–118.
- Boers, T.M. and J.B. Asher. 1982. A review of rainwater harvesting. *Agr. Water Mgt.* 5:145–158.
- Canadian Council of Ministers of the Environment. 2011. Canadian environmental quality guidelines summary table. 25 Aug. 2012. <<http://st-ts.ccme.ca/?lang=en&factsheet=229>>.
- Canadian Mortgage and Housing Corp. 2012. CMHC releases comprehensive report on housing in Canada. 1 Nov. 2012. <<http://www.cmhc-schl.gc.ca/en/corp/nero/nere/2008/2008-11-13-0815.cfm>>.
- Davies, P.J., I.A. Wright, O.J. Jonasson, and S.J. Findlay. 2010. Impact of concrete and PVC pipes on urban water chemistry. *Urban Water J.* 7:233–241.
- Despins, C., K. Farahbakhsh, and C. Leidl. 2009. Assessment of rainwater quality from rainwater harvesting systems in Ontario, Canada. *J. Water Supply* 58:117–134.
- Environment Canada. 2012a. Canadian climate normals 1971–2000. 11 Feb. 2012. <[http://www.climate.weatheroffice.gc.ca/climate\\_normals/results\\_e.html?stnID=5420&lang=c&dCode=0&StationName=MONTREAL&SearchType=Contains&province=ALL&provBut=&month1=0&month2=12](http://www.climate.weatheroffice.gc.ca/climate_normals/results_e.html?stnID=5420&lang=c&dCode=0&StationName=MONTREAL&SearchType=Contains&province=ALL&provBut=&month1=0&month2=12)>.
- Environment Canada. 2012b. National climate data and information archive. 26 July 2012. <[http://www.climate.weatheroffice.gc.ca/climateData/dailydata\\_e.html?timeframe=2&Prov=QC&StationID=10873&dlyRange=1993-08-26|2011-07-13&Year=2011&Month=7&Day=01](http://www.climate.weatheroffice.gc.ca/climateData/dailydata_e.html?timeframe=2&Prov=QC&StationID=10873&dlyRange=1993-08-26|2011-07-13&Year=2011&Month=7&Day=01)>.
- Farahbakhsh, K., C. Despins, and C. Leidl. 2009. Developing capacity for large-scale rainwater harvesting in Canada. *Water Qual. Res. J. Canada* 44:92–102.
- Forster, J. 1996. Patterns of roof runoff contamination and their potential implications on practice and regulation of treatment and local infiltration. *Water Sci. Technol.* 38:39–48.
- Fox, P., J. Rockstorm, and J. Barron. 2005. Risk analysis and economic viability of water harvesting for supplemental irrigation in semi-arid Burkina Faso and Kenya. *Agr. Syst.* 83:231–250.
- Ghisi, E. 2006. Potential for potable water savings by using rainwater in the residential sector of Brazil. *Build. Environ.* 41:1544–1550.
- Hart, C. and D. White. 2006. Water quality and construction materials in rainwater catchments across Alaska. *J. Environ. Eng. Sci.* 5:S19–S25.
- Hermann, T. and U. Schmida. 1999. Rainwater utilization in Germany: Efficiency, dimensioning, hydraulic and environment aspects. *Urban Water* 1: 307–316.
- Khastagir, A. and N. Jayasuriya. 2010. Optimal sizing of rainwater tanks for domestic water conservation. *J. Hydrol. (Amst.)* 381:181–188.
- Kumar, D.M. 2004. Roof water harvesting for domestic water security: Who gains and who loses? *Water Intl.* 29:43–53.
- Leopold, L.B. 1968. Hydrology for urban planning: A guidebook on the hydrologic effects of urban land use. U.S. Geological Survey Circ., U.S. Dept. Interior, Washington, DC.
- Roebuck, R.M. and R.M. Ashley. 2006. Predicting the hydraulic and life-cycle cost performance of rainwater harvesting systems using a computer based modelling tool. 7th Intl. Conf. Urban Drainage Modelling. p. 546–553.
- Save the Rain Campaign. 2010. Rainwater harvesting FAQ's. 11 Feb. 2012. <<http://www.savetherain.info/media-centre/rainwater-harvesting-faqs.aspx>>.
- United Nations Environment Programme. 2012. Examples of rainwater harvesting and utilisation around the world. 11 June 2012. <<http://www.unep.or.jp/ietc/publications/urban/urbanenv-2/9.asp>>.
- Van Metre, P.C. and B.J. Mahler. 2003. The contribution of particles washed from rooftops to contaminant loading to urban streams. *Chemosphere* 52:1727–1741.

Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

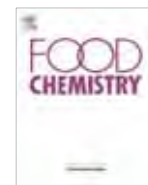
<http://www.elsevier.com/authorsrights>





Contents lists available at SciVerse ScienceDirect

Food Chemistry

journal homepage: [www.elsevier.com/locate/foodchem](http://www.elsevier.com/locate/foodchem)

## Fatty acid profiling of the seed oils of some varieties of field peas (*Pisum sativum*) by RP-LC/ESI-MS/MS: Towards the development of an oilseed pea

Manuel Ivan Villalobos Solis<sup>a</sup>, Anil Patel<sup>a</sup>, Valérie Orsat<sup>a</sup>, Jaswinder Singh<sup>b</sup>, Mark Lefsrud<sup>a,\*</sup>

<sup>a</sup> Bioresource Engineering Department, McGill University, Macdonald Campus, 21,111 Lakeshore Road, Ste-Anne-de-Bellevue, Quebec, Canada H9X 3V9

<sup>b</sup> Plant Science Department, McGill University, Macdonald Campus, 21,111 Lakeshore Road, Ste-Anne-de-Bellevue, Quebec, Canada H9X 3V9

### ARTICLE INFO

#### Article history:

Received 7 August 2012

Received in revised form 30 November 2012

Accepted 21 December 2012

Available online 16 January 2013

#### Keywords:

*Pisum sativum*

Seed oil

RP-LC/ESI-MS/MS

Fatty acids

### ABSTRACT

Reversed-phase liquid chromatography coupled to negative-ion electrospray tandem mass spectrometry (RP-LC/ESI-MS/MS) was used to study the fatty acid profile from the oil of harvested field pea (*Pisum sativum*) varieties as part of a research project to develop this legume as a commercial oilseed for Canada. The seed oils from pea samples contained palmitic and stearic acids as major saturated fatty acids. Oleic, linoleic and linolenic acids were the major unsaturated fatty acids found. Small percentages of other long chain fatty acids were also detected. This profile suggests that the species of field pea investigated might have the potential to be used as raw materials to develop a future new oilseed crop for the food industry. Fatty acid extracts did not require further manipulation before final analysis by RP-LC/ESI-MS/MS, indicating the utility and relative simplicity of this technique for future screening studies.

© 2013 Elsevier Ltd. All rights reserved.

### 1. Introduction

The field pea (*Pisum sativum* L.) also known as common pea, dry pea, green pea, yellow pea and garden pea (Ratnayake, Hoover, Shahidi, Perera, & Jane, 2001) is one of the major food legumes in the world and ranks fourth in the world production next to soybeans, peanuts and dry beans (Yoshida, Tomiyama, Tanaka, & Mizushima, 2007). Canada is the world largest producer and exporter of field peas. In 2010 of the 2.86 metric tonnes produced in Canada, approximately 65% was in Saskatchewan while Alberta and Manitoba produced 33% and 2% respectively (Pulse Canada. Canadian pulse industry: Situation, 2012).

This crop has an optimum growth temperature range of 12–18 °C and a cold tolerance down to –7 °C. Field pea protein content has been calculated at 23% (Thiessen, 2004; Wang & Daun, 2004) and as high as 25% (Singh, Kaur, Rana, & Sharma, 2010) and its carbohydrate content falls in the 50–60% range (Nikolopoulou, Grigorakis, Stasini, Alexis, & Iliadis, 2007). It is underutilised as whole or split in soups and stews, as hulls in high fibre bread, in the production of adhesives and carbonless paper (pea starch) and as an alternative protein source to soybean (*Glycine max*) and canola (*Brassica napus*) meal (Ratnayake et al., 2001). Published literature has shown low average oil content in field peas, ranging from 1.5% to 3.7% (Welch & Wynne Griffiths, 1984; Yoshida

et al., 2007) suggesting that current varieties have limited use as an oil crop.

A recent study conducted at the Bioresource Engineering Department of McGill University screened the lipid content of 198 varieties of field pea and found 30 candidates with mean contents in the range of 1.15–6.31%. With further research, these oil yields could allow the development of the field pea as an oilseed with a commercial lipid content greater than 20%, and potentially compete with the primary oilseeds grown in Canada, namely, canola (with >35% oil) and soybean (>20% oil). To achieve this long-term objective, the development of a new variety of field pea requires two major steps: screening for lipid content and its improvement through breeding.

Benefits of a new oilseed feedstock for Canada include cold weather tolerance with similar annual grain yields as soybean, nitrogen fixation (with nitrogen storage of about 65–70%) that could allow a reduction in the costs of fertilisers, and a source of high quality protein.

A field pea variety rich in poly-unsaturated fatty acids ( $\omega$ -3,  $\omega$ -6), could become a dietary source contributing to cardiovascular health, physiological functions and maintenance (Ryan, Galvin, O'Connor, Maguire, & O'Brien, 2007). Furthermore, field pea as an alternative oilseed holds potential as a bio-oil resource for the production of oleochemicals derived from it. One example is commercial biodiesel. By 2011 the government of Canada targeted that 2% of the diesel fuel and heating oil needed to be from renewable sources, where an oilseed pea variety could be developed to compete with canola and soybean. Other industrial applications include lubricants produced from high oxidative stability oils

\* Corresponding author. Tel.: +1 514 398 7967; fax: +1 514 398 8387.

E-mail addresses: [manuel.villalobosolis@mail.mcgill.ca](mailto:manuel.villalobosolis@mail.mcgill.ca) (M.I. Villalobos Solis), [anil.patel@mail.mcgill.ca](mailto:anil.patel@mail.mcgill.ca) (A. Patel), [valerie.orsat@mcgill.ca](mailto:valerie.orsat@mcgill.ca) (V. Orsat), [jaswinder.singh@mcgill.ca](mailto:jaswinder.singh@mcgill.ca) (J. Singh), [mark.lefsrud@mcgill.ca](mailto:mark.lefsrud@mcgill.ca) (M. Lefsrud).



(e.g. high oleic acid content) or coating applications such as paints, inks and varnishes from low oxidative stability oils (e.g. high linolenic acid content) (Cahoon, 2003).

Fundamental to screening varieties of field peas for high oil content is an appropriate method for the characterisation of fatty acids. Fatty acid profiling can give information about suitable candidates for breeding and selection experiments, classification of newly produced varieties and information on appropriate strains for different applications, be it oleochemicals production or human consumption.

However, the analysis of lipids is quite complicated because of the high degree of complexity and the heterogeneity of their components. Especially from vegetable samples, the number of lipid molecules is higher than that of animal or prokaryotic organisms hence introducing more degree of diversity (Seiwert, Giavalisco, & Willmitzer, 2009). At present, there is no single analytical tool capable of simultaneously identifying and quantifying all these lipid species without combining various technologies.

Gas chromatography coupled to flame ionisation (GC–FID) has been the method of choice for fatty acid screening studies but it requires sample derivatisation which is time-consuming and may compromise the stability of the molecules with undesirable alterations such as isomerisation, oxidation and thermal degradations (Nishiyama-Naruke, Souza, Carnelos, & Curi, 1998); for these reasons, more practical alternatives have been explored in recent years which include the use of liquid chromatography coupled to mass spectrometry (LC–MS). This technique also allows fatty acids with the same equivalent carbon number (ECN) to be discriminated by their masses or fragments (Byrdwell, 2005, chap. 5), a task that is relatively more difficult to achieve with traditional two dimensional detectors like refractive index, UV/Vis and evaporative light scattering detector (ELSD).

According to Lima and Abdalla (2002), some advantages that electrospray ionisation in the LC–MS analysis of fatty acids offers are improved separation of these molecules from the complex biological matrices; the possibility of using high flows (100–1000  $\mu\text{L min}^{-1}$ ) usually necessary in HPLC separations without the loss of sensitivity; and the omission of chemical derivatisations, like esterification or transesterification prior to analysis, which in turns speeds the assay, reducing it to mainly two steps: saponification and extraction of fatty acids (Kurata, Yamaguchi, & Nagai, 2005).

By employing electrospray ionisation in the negative mode, ions of fatty acids can be monitored as deprotonated molecules ( $[M-H]^-$ ) with the aid of acetic acid, formic acid or ammonium acetate which are added to the mobile phases in order to facilitate the ionisation (Kerwin, Wiens, & Ericsson, 1996; Perret, Gentili, Marchese, Sergi, & Caporossi, 2004). By conducting tandem mass spectrometry, it is possible to determine branched points or double bonds in the fatty acid chains, providing a more accurate detection method (Lima & Abdalla, 2002).

The aim of the present work was to develop an analytical method for the identification of fatty acids in pea oil samples based on reverse phase high performance liquid chromatography interfaced to negative ion electrospray tandem mass spectrometry (RP-LC/ESI-MS/MS) in order to help in the long term development of field pea as an oilseed.

Oil extraction was performed by a modified version of the hexane: isopropanol method proposed by Ryan, Galvin, O'Connor, Maguire, and O'Brien (2007) which according to previous experiments in our laboratory had the highest yields of pea oil (unpublished data). A mild alkaline hydrolysis procedure at low temperature was also employed in order to release the fatty acid molecules from complex lipid structures.

The analytical methodology proposed here may also prove useful for the separation and analysis of fatty acids from other seed oil sources as it was also tested for samples of soybean and canola.

## 2. Materials and methods

### 2.1. Biological samples

Six varieties of field pea (*P. sativum*) catalogued as 29579, 29595, 29610, 45760, Big Pea and Galena were analysed in this study (Plant Gene Resources of Canada, Saskatoon, SK; U.S. Department of Agriculture, Pullman, WA). Seeds were harvested in the year 2010 from either Lefsrud's farm (Viking, AB, Canada) and/or Macdonald Campus of McGill University (Ste-Anne-de-Bellevue, QC, Canada). Pods of the samples were previously dried in an oven at 60 °C for 48 h with no further treatment. Seeds from cultivars of soybean (*G. max*), OAC Champion variety, and canola (*B. napus*), Roper variety, were also screened for fatty acid profile and these were grown and harvested at Lefsrud's farm (Viking, AB, Canada) and Agrocentre Belcan (Sainte-Marthe, QC, Canada) respectively.

### 2.2. Chemicals and reagents

For sample preparation the following analytical grade solvents and reagents were used: 2-propanol LC–MS Chromasolv<sup>®</sup> was purchased from Sigma–Aldrich (St. Louis, MO, USA); *n*-hexane 95% optima; chloroform HPLC-grade 99.8; petroleum ether certified ACS; hydrochloric acid reagent A.C.S.; sodium sulphate anhydrous granular, 10–60 Mesh, Certified ACS and potassium hydroxide, extra pure, ca. 85% flakes, were all obtained from Fisher Scientific (Fair Lawn, NJ, USA). For LC–MS analysis, acetic acid was acquired from Sigma–Aldrich (St. Louis, MO, USA); Optima<sup>®</sup> LC–MS grade acetonitrile, methanol and water were purchased from Fisher Scientific (Fair Lawn, NJ, USA). All solvents were used without further purification.

The saturated fatty acid standards, myristic ( $C_{14:0}$ ), palmitic ( $C_{16:0}$ ), stearic ( $C_{18:0}$ ), arachidic ( $C_{20:0}$ ), behenic ( $C_{22:0}$ ) and lignoceric ( $C_{24:0}$ ); as well as the mono- and polyunsaturated standards, palmitoleic ( $C_{16:1}$ ), oleic ( $C_{18:1}$ ), linoleic ( $C_{18:2}$ ),  $\alpha$ - and  $\gamma$ -linolenic ( $C_{18:3}$ ), gadoleic ( $C_{20:1}$ ), eicosadienoic ( $C_{20:2}$ ), erucic ( $C_{22:1}$ ) and nervonic ( $C_{24:1}$ ) were purchased from Nu-Check Prep (Elysian, MN, USA). Individual stock solutions containing 17  $\mu\text{g}/\mu\text{L}$  of each standard were prepared by dissolving the analytes in HPLC-grade chloroform; composite working standard solutions were prepared by mixing adequate volumes of diluted stock solutions and adding acetonitrile as needed. All solutions were flushed with nitrogen every time after use and stored at –80 °C.

### 2.3. Sample preparation

#### 2.3.1. Lipid extraction and oil determination

Oil extraction from the different biological samples was achieved by a modified version of the method described by Ryan, Galvin, O'Connor, Maguire, and O'Brien (2007). Eight grams of ground samples were placed in 50 ml Teflon-lined screw-capped glass centrifuge tubes of known mass in three biological replicates. A fourth tube with the same amount of ground sample was prepared and used as a control in order to measure moisture content.

Oil was extracted with 12 ml of hexane–isopropanol (3:2, v/v) at room temperature and constant stirring for 1 h. The samples were then vortex for 30 s followed by centrifugation at 3000g for 10 min. The solvent layers were collected in 50 ml glass centrifuge tubes. The pellets were washed two more times with 10 ml of the organic solvent, each time vortexing for 30 s, centrifuging at 3000g for 10 min and recovering the hexane: isopropanol layers. The washed pellets were dried under nitrogen for up to 30 min at 70 °C until the remaining solvent was completely evaporated. The tubes were un-capped and placed in the oven for 24 h at 95 °C along with the control tubes for each sample. After the drying

was over and the tubes were allowed to reach room temperature their final masses were measured and recorded. The difference between the initial and final mass of the control tube, which represents the moisture loss during the drying period, was subtracted from the difference in mass of each sample tube in order to calculate the oil content percentage of the samples.

The recovered hexane–isopropanol extracts were centrifuged at 3000g for 10 min and the supernatant transferred to a new known mass glass centrifuge tube. The solvent was nitrogen evaporated under 50 °C and the pure oil was redissolved in chloroform for storage and further sample preparation.

### 2.3.2. Fatty acids extraction by saponification

After solvent evaporation, approximately 100 mg of extracted oil from each sample were treated with 5 ml of methanolic KOH (0.5 M) at 55 °C for 20 min. When the saponification reaction was ready, 2.5 ml of concentrated HCl were added. The mixture was then transferred to a 30 ml separatory funnel and the fatty acids extracted twice with 5 ml of petroleum ether.

### 2.3.3. Removal of water and non-lipid contents

The combined petroleum ether extracts from the saponification reaction were transferred into glassware funnels containing 15 g of anhydrous sodium sulphate previously rinsed with 5 ml of petroleum ether. The filtrates were collected in 50 ml glass centrifuge tubes. Additional washes of 2 ml petroleum ether were done and collected in the same tubes. Nitrogen was gently blown through the combined washes at 40 °C. When most of the solvent was evaporated from the solutions, the samples were re-dissolved with 10 ml of HPLC grade chloroform, flushed with nitrogen and stored at –80 °C. From these stock solutions 3 ml were diluted each time with 7 ml of LC–MS grade acetonitrile for further RP-LC/ESI-MS/MS analysis.

## 2.4. RP-LC/ESI-MS/MS analysis

### 2.4.1. RP-LC system

The HPLC system was a Surveyor Plus™ LC equipped with an analytical pump, an autosampler with a 25 µL sample loop and a UV/Vis detector (Thermo Electron, San Jose, CA, USA). A reverse phase Hypersil Gold C18 Column® with a particle size of 5 µm (100 × 4.6 mm, i.d. × length) (Thermo Fisher Scientific, Waltham, MA, USA) was used to analyse the fatty acid fractions collected. A ternary solvent gradient elution method consisting of 25% acetonitrile–75% water (mobile phase A), 100% acetonitrile (mobile phase B) and 100% methanol (mobile phase C) all containing 0.12% acetic acid was programmed to separate the molecules. The mobile phases changed from 46.00% A, 54.00% B and 0.00% C at time zero to 12.00% A, 85.00% B and 3.00% C in 17 min; from the last percentages to 12.00% A, 0.00% B and 88.00% C in 9 min and then to 0.00% A, 0.00% B and 100.00% C in 1 min. This gradient remained constant for 7 min, then returned to the initial composition in 1 min and maintained for 10 additional minutes. The flow rate employed at all times was 1 ml min<sup>-1</sup>. The column was kept at a constant temperature of 35 °C. 10 µL of each biological replicate were injected three times (technical replicates) into the column for analysis. The effluent from the LC system was split 1:10 before introduction to the ESI source.

### 2.4.2. Tandem mass spectrometry

A LTQ-XL 2D linear ion trap mass spectrometer (Thermo Electron, San Jose, CA, USA) equipped with an ESI source was coupled on-line to the RP-LC system described before. The spectrometer was operated in the negative ion (NI) mode. The source voltage was 3.50 kV. Capillary temperature was maintained at 350.00 °C. The settings for the sheath and auxiliary gas flows were respec-

tively 30.00 and 5.00 instrument units. Data acquisition and processing were performed using Xcalibur® version 2.0.7 software (Thermo Electron, San Jose, CA, USA).

A data dependent tandem mass spectrometric (MS/MS) experiment consisting of two scan events was programmed on the data system in order to identify the fatty acid molecules in the extracts. The full scan mode was chosen as the acquisition mode. The mass spectra were obtained between *m/z* 200 and 800. During MS/MS, helium was used as the collision gas for collision induced dissociation (CID). The collision energy was set at 35.00 eV and the isolation width set at 2.0 u. The system was calibrated by infusion of a normal mass range calibration solution containing caffeine, L-methionyl–arginyl–phenylalanyl–alanine acetate salt (MRFA) and Ultramark 1621 at a flow rate of 5 µL/min.

## 2.5. Data evaluation

Fatty acids were identified based on elution patterns from standard working solutions and characteristic parent peaks detected in the mass spectra of each compound. Structure elucidation and confirmation were conducted taking as reference the MS/MS spectra of each standard. The fatty acid composition was calculated as a relative percentage of the total area of all peaks.

## 3. Results and discussion

### 3.1. Fatty acid composition of pea oil

The obtained yields of pea oil (from 1.50% for variety 29579 to 2.03% for variety Galena), soybean oil (15.32%) and canola oil (36.10%) are in close agreement with what literature has reported for these crops (Hammond, Johnson, Su, Wang, & White, 2005; Ryan, Galvin, O'Connor, Maguire, & O'Brien, 2007; Shahidi, 1990, chap. 6; Wang & Daun, 2004; Welch & Wynne Griffiths, 1984; Yoshida et al., 2007). According to Yoshida et al. (2007) predominant lipid components of field pea oil are phospholipids (PLs) and triacylglycerols (TAGs). These authors analysed four varieties of field pea and found that between both lipid classes the amount of PLs (55.20–61.30 wt.%) was higher than that of TAGs (31.20–40.30 wt.%) mainly because peas are not oilseeds, but typical vegetable seeds. The same researchers also reported minor quantities of free fatty acids (1.3–2.7 wt.%), 1,2-diacylglycerols (1.0–2.2 wt.%), 1,3-diacylglycerols (1.0–1.8 wt.%), steryl esters (0.8–2.4 wt.%) and hydrocarbons (0.5–0.9 wt.%) for the same varieties.

The previous data suggests that the bulk of the fatty acid fraction of pea oil is bound to other molecules and this is the reason of why a saponification procedure was employed here.

The fatty acid profile of the different varieties of field pea, canola and soybean samples analysed by RP-LC/ESI-MS/MS and their relative percentage in oil are presented in Table 1. Linoleic acid (C<sub>18:2</sub>) was the major component in field peas and soybean oil; whereas oleic acid (C<sub>18:1</sub>) occupied the highest percentage in canola oil; all these results agree with published data (El-Saied, Amer, & Gabran, 1981; Hammond et al., 2005; Murcia & Rincon, 1992; Nishiyama-Naruke et al., 1998; Ryan, Galvin, O'Connor, Maguire, & O'Brien, 2007; Shahidi, 1990, chap. 6; Welch & Wynne Griffiths, 1984; Yoshida et al., 2007).

The levels of saturated fatty acids (SFAs) in all samples ranged from 10.83% for Canola Roper to 20.13% for pea Galena variety. Total mono-unsaturated fatty acids (MUFAs) ranged from 26.60% for pea variety 45760 to 58.15% for Canola Roper. Poly-unsaturated fatty acids (PUFAs) ranged from 30.92% for Canola Roper to 56.80% for pea variety 29610. The most abundant PUFA in all samples was linoleic acid, MUFA was oleic acid and SFA was palmitic acid (C<sub>16:0</sub>). Unknown molecules were also seen in the different

**Table 1**

Fatty acid composition of the oilseeds analysed. The peak area of each component is reported as a percent of the total area of all peaks. Values are the means of three biological replicates  $\pm$  standard deviation.

Fatty acids or <i>m/z</i> values detected	Relative peak area (% of fatty acids in extracted oil)							
	29579	29595	Big Pea	29610	45760	Galena	OAC champion	Roper
227 <sup>a</sup>	0.09 $\pm$ 0.02	ND	ND	ND	ND	ND	ND	ND
C <sub>16:0</sub>	9.01 $\pm$ 0.11	7.54 $\pm$ 0.52	7.63 $\pm$ 0.64	7.17 $\pm$ 0.23	8.91 $\pm$ 0.52	7.20 $\pm$ 0.32	8.77 $\pm$ 0.18	3.67 $\pm$ 0.30
C <sub>18:0</sub>	5.12 $\pm$ 0.24	5.48 $\pm$ 0.31	5.02 $\pm$ 0.15	4.89 $\pm$ 0.14	6.27 $\pm$ 0.75	5.94 $\pm$ 0.49	7.31 $\pm$ 0.73	3.44 $\pm$ 0.25
C <sub>18:1</sub>	26.9 $\pm$ 0.07	35.79 $\pm$ 1.35	27.87 $\pm$ 0.74	26.56 $\pm$ 1.94	25.52 $\pm$ 0.80	34.51 $\pm$ 0.91	27.39 $\pm$ 0.63	54.90 $\pm$ 3.16
C <sub>18:2</sub>	39.45 $\pm$ 1.34	35.95 $\pm$ 2.22	42.32 $\pm$ 3.55	44.78 $\pm$ 0.89	43.98 $\pm$ 2.22	35.85 $\pm$ 3.64	43.58 $\pm$ 1.33	21.38 $\pm$ 0.91
C <sub>18:3</sub>	14.01 $\pm$ 0.60	9.59 $\pm$ 0.72	11.25 $\pm$ 1.28	12.13 $\pm$ 0.58	9.89 $\pm$ 0.59	6.01 $\pm$ 0.26	8.97 $\pm$ 0.64	9.54 $\pm$ 1.36
C <sub>20:0</sub>	1.29 $\pm$ 0.10	1.17 $\pm$ 0.17	1.16 $\pm$ 0.29	1.09 $\pm$ 0.19	1.30 $\pm$ 0.10	1.90 $\pm$ 0.18	0.85 $\pm$ 0.14	1.32 $\pm$ 0.10
C <sub>20:1</sub>	0.76 $\pm$ 0.09	0.84 $\pm$ 0.37	0.91 $\pm$ 0.14	0.72 $\pm$ 0.09	0.90 $\pm$ 0.12	1.50 $\pm$ 0.27	0.30 $\pm$ 0.03	2.76 $\pm$ 0.23
307 <sup>a</sup>	ND	ND	0.14 $\pm$ 0.01	ND	ND	ND	ND	ND
C <sub>22:0</sub>	0.57 $\pm$ 0.09	ND	0.46 $\pm$ 0.14	0.51 $\pm$ 0.34	0.51 $\pm$ 0.10	1.20 $\pm$ 0.35	1.67 $\pm$ 0.25	1.27 $\pm$ 0.21
C <sub>22:1</sub>	0.35 $\pm$ 0.04	0.50 $\pm$ 0.19	0.29 $\pm$ 0.09	0.25 $\pm$ 0.04	0.18 $\pm$ 0.02	0.37 $\pm$ 0.06	ND	ND
C <sub>24:0</sub>	1.78 $\pm$ 0.10	2.73 $\pm$ 0.70	1.90 $\pm$ 0.72	1.26 $\pm$ 0.35	1.55 $\pm$ 0.14	3.88 $\pm$ 1.19	1.07 $\pm$ 0.27	1.12 $\pm$ 0.19
C <sub>24:1</sub>	ND	ND	ND	ND	ND	ND	ND	0.50 $\pm$ 0.16
325 <sup>a</sup>	ND	ND	ND	ND	0.08 $\pm$ 0.01	0.15 $\pm$ 0.01	ND	ND
353 <sup>a</sup>	ND	ND	ND	ND	0.16 $\pm$ 0.04	0.35 $\pm$ 0.08	ND	ND
381 <sup>a</sup>	0.24 $\pm$ 0.03	ND	0.35 $\pm$ 0.14	0.26 $\pm$ 0.08	0.26 $\pm$ 0.04	0.55 $\pm$ 0.14	0.03 $\pm$ 0.01	ND
395 <sup>a</sup>	0.34 $\pm$ 0.02	0.42 $\pm$ 0.03	0.40 $\pm$ 0.09	0.24 $\pm$ 0.05	0.55 $\pm$ 0.03	0.57 $\pm$ 0.10	0.07 $\pm$ 0.01	0.10 $\pm$ 0.01
SFAs	17.86 $\pm$ 0.64	16.92 $\pm$ 1.47	16.17 $\pm$ 1.89	14.92 $\pm$ 0.88	18.54 $\pm$ 1.39	20.13 $\pm$ 2.40	19.67 $\pm$ 1.48	10.83 $\pm$ 0.90
MUFAs	28.10 $\pm$ 0.08	37.29 $\pm$ 0.86	29.08 $\pm$ 0.69	27.54 $\pm$ 1.86	26.60 $\pm$ 0.70	36.39 $\pm$ 0.96	27.69 $\pm$ 0.65	58.12 $\pm$ 2.85
PUFAs	53.46 $\pm$ 0.76	44.68 $\pm$ 1.59	53.51 $\pm$ 2.28	56.80 $\pm$ 1.17	53.87 $\pm$ 2.11	41.86 $\pm$ 3.42	52.55 $\pm$ 1.96	30.92 $\pm$ 2.23
Unknown	0.58 $\pm$ 0.05	0.42 $\pm$ 0.03	1.19 $\pm$ 0.33	0.63 $\pm$ 0.17	1.05 $\pm$ 0.07	1.62 $\pm$ 0.32	0.09 $\pm$ 0.02	0.10 $\pm$ 0.01

ND, not detected; SFAs, Saturated fatty acids; MUFAs, mono-unsaturated fatty acids; PUFAs, poly-unsaturated fatty acids.

<sup>a</sup> Undefined *m/z* value detected.

**Table 2**

Comparative table of the fatty acid composition of pea oil (%). Values are mean determinations.

Fatty acid	Mean values found (%)	Reported values (%)			
		El-Saied et al. (1981) <sup>a</sup>	Murcia and Rincon (1992) <sup>b</sup>	Wang and Daun (2004) <sup>a</sup>	Ryan, Galvin, O'Connor, Maguire, and O'Brien (2007) <sup>a</sup>
Palmitic (C <sub>16:0</sub> )	6.76	12.79	16.40	10.65	10.65
Stearic (C <sub>18:0</sub> )	2.86	2.41	15.20	3.29	3.29
Oleic (C <sub>18:1</sub> )	31.04	14.67	23.50	23.22	28.15
Linoleic (C <sub>18:2</sub> )	46.06	53.99	32.90	45.63	47.59
Linolenic (C <sub>18:3</sub> )	11.12	9.04	12.00	13.69	9.29
Arachidic (C <sub>20:0</sub> )	0.13	NR	NR	0.79	0.22
Gadoleic (C <sub>20:1</sub> )	0.25	NR	NR	0.62	0.21
Erucic (C <sub>22:1</sub> )	0.03	NR	NR	0.24	ND
Lignoceric (C <sub>24:0</sub> )	1.77	NR	NR	0.33	NR

NR – not reported, ND – not detected.

<sup>a</sup> Determined by GC–FID.

<sup>b</sup> Determined by GC–MS.

samples, but their relative percentages were low (less than 1.50%), a discussion about their possible identities is presented in the mass spectrometry analysis section of this paper.

Palmitic acid was detected in all field pea varieties but found at the highest level in 29579 (9.01%) than in other species. Stearic acid (C<sub>18:0</sub>) was higher in field pea seed from variety 45760 (6.27%) than in all others. Arachidic acid (C<sub>20:0</sub>) was present in all pea samples at relative similar percentages from 1.09% for variety 29610 to 1.90% for Galena. Behenic acid (C<sub>22:0</sub>) occurred in field pea samples at less than 1.50%, with Galena variety having the highest percentage (1.20%) and Big Pea variety the lowest (0.46%). According to Balogun and Fetuga (1985) oils with high levels of behenic acid may be difficult to digest by humans and animals, thus highlighting the value that pea oil could have for food purposes. Other SFA found in this study was lignoceric acid (C<sub>24:0</sub>) a long chain fatty acid that has not been reported extensively for field pea oil but has been found in other members of the Leguminosae family (Bağcı & Şahin, 2004).

The major MUFA found in all samples was oleic acid. Specifically for field peas, seeds from varieties 29595 (35.79%) and Galena (34.51%) had the highest levels of it. Other acids found and belonging to this category were gadoleic (C<sub>20:1</sub>) and erucic (C<sub>22:1</sub>) acids.

The first of them had the highest level, within varieties of field pea, in oil from variety Galena (1.50%), whereas erucic acid presented its highest level in variety 29595. Nervonic acid (C<sub>24:1</sub>) was not present in field peas neither in soybean, but canola had a level of 0.50%.

The isomer  $\alpha$ -linolenic acid (C<sub>18:3</sub>) was found in all samples but its levels were lower than oleic and linolenic acid as stated before. Studies by Hilditch and Williams (1964) suggested that linolenic acid is either absent or present in very small amounts in most Leguminosae seed oils; however, the percentages calculated here, resulted in the third largest in importance as opposed to the trace amounts (<0.05%) found of arachidic, gadoleic and lignoceric acids.

A comparative table (Table 2) of the fatty acid composition obtained in this study for pea oil to what other authors have found by gas chromatography–flame ionisation detector (GC–FID) or gas chromatography–mass spectrometry (GC–MS) is presented.

According to Murcia and Rincon (1992) fatty acid composition of pea oil is influenced by agro-climatic conditions as it is known that the enzymes in fatty acid biosynthesis depend on these factors. The reduced quantity of palmitic and linolenic acids found in peas in comparison to the levels of linoleic acid can be explained by the natural growth of these legumes as it has been observed

that, as the diameter of peas increases during growth period, the proportion of these fatty acids diminishes. Polyunsaturated fatty acids in pea oil are commonly found in seed oils, and they play a role in the initial oxidative processes whereby fatty acids are converted to carbohydrates, because of their ease of peroxidation (Holman, 1981).

Our results indicate that due to the content of PUFAs, pea oil can be advisable for human consumption. Oleic acid has been reported as effective in reducing cardiovascular risk; linoleic acid, which was the major component of pea oil, is considered an “essential” fatty acid for human growth, physiological functions and maintenance; and  $\alpha$ -linolenic acid is regarded as potentially functional in reducing cardiovascular risk through its particular omega-3 structure (Shahidi, 1990, chap. 6). The nutritional value of pea oil is also shared with soybean and canola as their fatty acid profile found and compared in published literature is also dominated by these PUFAs. As an example, it was found that for canola oil the proportion of  $\alpha$ -linolenic acid relative to linoleic acid was of approximate 2.24:1, which is close to the ideal 2:1 human consumption proportion that has been reported for this crop (Shahidi, 1990, chap. 6).

### 3.2. Mass spectrometry analysis

The full scan spectra in the negative ion mode of all fatty acid analytes were dominated by the  $[M-H]^-$  anions as base peaks and for that reason these were taken as precursor ions in order to conduct further fragmentation studies by tandem mass spectra. Table 3 shows the masses of the base peaks observed for ESI in the negative mode.

Data dependent experiments were programmed on the Xcalibur<sup>®</sup> analytical software with the main objective of conducting MS/MS fragmentation of standards and acquire all the possible characteristic product spectra for subsequent structure elucidation and confirmation of analytes in oil samples. Because many compounds can have the same molecular mass, the intact mass of a compound is not a unique identifier and thus ion transitions, in which a parent mass gives a set of particular fragment peaks, can be monitored to decrease the likelihood of false positives results.

The CID MS/MS spectra of saturated fatty acids were dominated by  $[M-H]^-$  peaks and the fragment ions  $[M-18]^-$  produced by a loss of H<sub>2</sub>O from the carboxyl group (Kerwin et al., 1996; Perret

et al., 2004). As stated in the literature, the NI MS/MS spectra of arachidic (C<sub>20:0</sub>) and behenic (C<sub>22:0</sub>) acids also yielded fragment ions at  $m/z$  183 which are explained by charge remote fragmentations (CRFs), a class of gas-phase decompositions that occur physically remote from the charge site and which are analytically useful in the determination of double bond positions and branching in the aliphatic chain (Kerwin et al., 1996; Perret et al., 2004). The spectra of mono-unsaturated standards was also dominated by  $[M-H]^-$  peaks and the fragment ions  $[M-18]^-$  although more fragments could be seen with respect to that of the saturated standards (Fig. 1). A study by Kerwin et al. (1996) showed that the MS/MS fragmentation of monounsaturated compounds in the NI mode of ESI produced an ion spectra very similar to those of saturated fatty acids; however, by limiting the scanning range of the most abundant ions of the spectrum they could increase the signal intensity and, in this way, determine more fragments that could be rationalised by the CRF process described before.

Examination of the  $\alpha$ -(18:3 $\omega$ 6) and  $\gamma$ -(18:3 $\omega$ 3) linolenic acid isomers showed that the most intense fragments peaks from the precursor  $[M-H]^-$  ions were formed by the loss of H<sub>2</sub>O ( $m/z$  259) and CO<sub>2</sub> ( $m/z$  233), this information agrees with published literature (Kerwin et al., 1996). The CID spectra of these poly-unsaturated fatty acids as well as of some mono-unsaturated fatty acids were also characterised by a series of ion peaks each one separated by 14  $m/z$  units, representing, according to Perret et al., 2004, CRFs cleavages of consecutive carbon bonds in the fatty acid chains and which were also useful to localise double bonds as the series were interrupted by gaps at their locations (Fig. 2).

As mentioned previously, some unknown peaks were also observed specifically with  $m/z$  values of 227, 307, 325, 353, 381 and 395. However, these values are in the range of most fatty acids  $m/z$  values found, so, some hypothesis were made about their identities.

For peaks with  $m/z$  227 and 307, it is believed that both are related to the mass of myristic (C<sub>14:0</sub>) and eicosadienoic (C<sub>20:2</sub>) acids respectively. Retention time data of these two unknown peaks in comparison to those of standards also suggested the presence of these acids in the oil of pea varieties 29579 and Big Pea (8.79 min  $\pm$  0.02 for  $m/z$  277 in 29579 compared to 8.71 min  $\pm$  0.02 to the myristic acid standard; and 12.85 min  $\pm$  0.90 for  $m/z$  307 compared to 13.86 min  $\pm$  0.04 for the eicosadienoic acid standard). Some authors have reported relative low percentage values of myristic and

**Table 3**

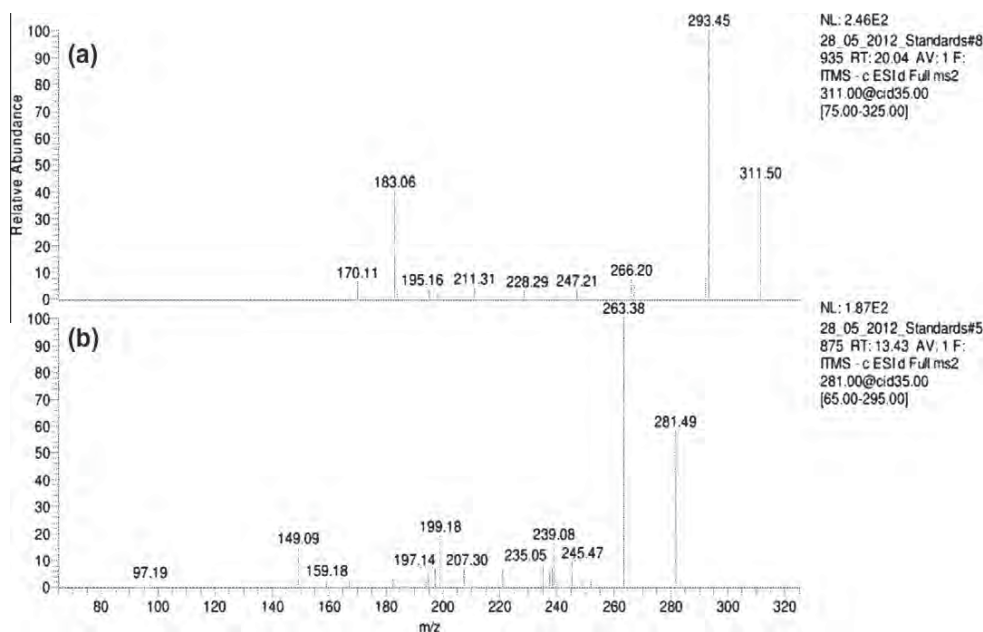
Exact molecular mass,  $[M-H]^-$  ions observed, retention times (RT) and limits of detection (LODs) and quantification (LOQs) for fatty acid standards analysed by RP-LC/ESI-MS.

Compound	Exact masses	$[M-H]^-$ ions detected <sup>a</sup>	RT (min) <sup>a</sup>	LODs (ng) <sup>b</sup>	LOQs (ng) <sup>b</sup>
<i>Saturated fatty acid standards</i>					
Myristic (C <sub>14:0</sub> )	228.20	227.20 $\pm$ 0.02	8.71 $\pm$ 0.02	24.06	80.20
Palmitic (C <sub>16:0</sub> )	256.24	255.25 $\pm$ 0.03	12.38 $\pm$ 0.04	26.41	88.06
Stearic (C <sub>18:0</sub> )	284.27	283.29 $\pm$ 0.02	16.28 $\pm$ 0.04	23.16	77.22
Arachidic (C <sub>20:0</sub> )	312.30	311.32 $\pm$ 0.02	19.77 $\pm$ 0.05	5.46	18.20
Behenic (C <sub>22:0</sub> )	340.33	339.38 $\pm$ 0.02	23.97 $\pm$ 0.10	28.50	95.01
Lignoceric (C <sub>24:0</sub> )	368.36	367.41 $\pm$ 0.02	29.56 $\pm$ 0.05	22.69	75.66
<i>Monounsaturated fatty acid standards</i>					
Palmitoleic (C <sub>16:1</sub> )	254.22	253.22 $\pm$ 0.03	9.41 $\pm$ 0.04	22.95	76.53
Oleic (C <sub>18:1</sub> )	282.25	281.26 $\pm$ 0.03	13.09 $\pm$ 0.05	19.55	65.17
Gadoleic (C <sub>20:1</sub> )	310.28	309.34 $\pm$ 0.03	17.02 $\pm$ 0.06	24.39	81.33
Erucic (C <sub>22:1</sub> )	338.31	337.37 $\pm$ 0.04	20.12 $\pm$ 0.07	2.77	9.26
Nervonic (C <sub>24:1</sub> )	366.34	365.38 $\pm$ 0.04	24.41 $\pm$ 0.08	9.82	32.75
<i>Polyunsaturated fatty acid standards</i>					
Linoleic (C <sub>18:2</sub> )	280.24	279.26 $\pm$ 0.03	10.31 $\pm$ 0.02	15.13	50.45
$\alpha$ -Linolenic (C <sub>18:3</sub> )	278.22	277.25 $\pm$ 0.03	8.11 $\pm$ 0.03	17.82	59.42
$\gamma$ Linolenic (C <sub>18:3</sub> )	278.22	277.27 $\pm$ 0.02	8.44 $\pm$ 0.02	22.99	76.65
Eicosadienoic (C <sub>20:2</sub> )	308.27	307.30 $\pm$ 0.01	13.86 $\pm$ 0.04	23.20	77.36

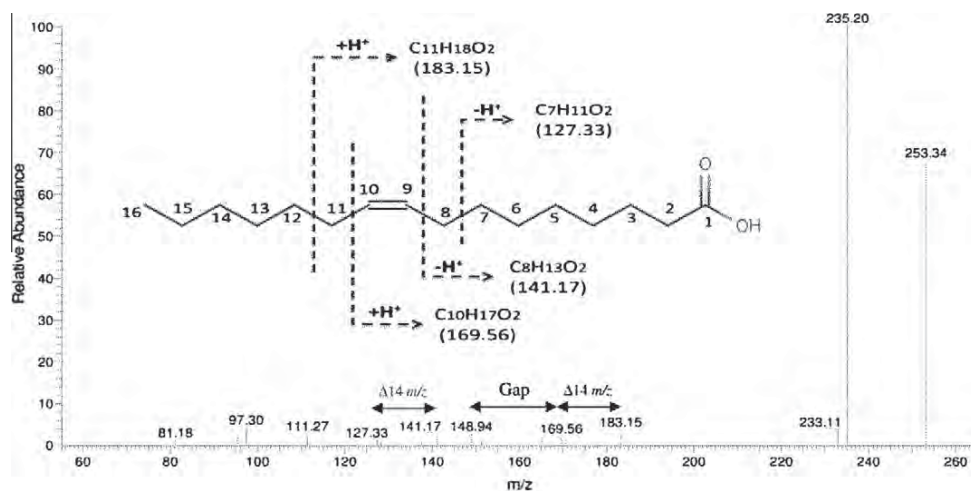
<sup>a</sup> Values presented are the means of five determinations  $\pm$  standard deviation.

<sup>b</sup> Values based on the equation: (SD/ $m$ ) \* 3.3 or 9. Where SD is the standard deviation of the response (peak areas) and  $m$  is the slope of each calibration curve at levels approximating the LOD or LOQ, respectively.





**Fig. 1.** Negative ion ESI-MS/MS of (a) arachidic ( $C_{20:0}$ ) and (b) oleic ( $C_{18:1}$ ) acids. The parent ions yielding each spectra are the  $[M-H]^-$  peaks detected in full scan experiments and these were also seen here as  $m/z$  311.50 and 281.49. For arachidic acid, peaks at  $m/z$  293.45 and at  $m/z$  183.05 can be explained by the loss of water from the carboxyl moiety and by CRFs processes respectively. For oleic acid, peak at  $m/z$  263.38 can be rationalised by the loss of water from the carboxyl moiety.



**Fig. 2.** NI ESI-MS/MS spectrum of palmitoleic acid ( $C_{16:1}$ ). Peak at  $m/z$  235.20 can be explained by loss of water from the carboxyl moiety and peak  $m/z$  253.34 is the  $[M-H]^-$  ion previously observed. An ion series with an inter-peak spacing of 14  $m/z$  units was also appreciated and can be rationalised by cleavage of consecutive C–C single bonds in the fatty acid chain as shown in the inset structure. The peak at  $m/z$  169.56 was used as an indicator of the presence of a double bond in the molecule.

eicosadienoic acids in different varieties of field pea supporting the present hypothesis (Coxon & Wright, 1985; El-Saied et al., 1981; Yoshida et al., 2007). However, for this study a solid conclusion cannot be given at this time, since the lack of MS/MS spectra for these two peaks did not allow the positive identification of the peaks as real fatty acids or false positives results.

Identity of peaks with  $m/z$  values close to 381 and 395 which were seen in some pea oil samples and in soybean and canola oil are believed to be related to pentacosylic ( $C_{25:0}$ ) and cerotic ( $C_{26:0}$ ) acids respectively. Since no standards of these acids were purchased and the MS/MS spectra of both peaks showed a similar pattern to those of saturated fatty acids, quick searches in the online available databases LIPID MAPS MS Prediction Tool ([\[www.lipidmaps.org/tools/\]\(http://www.lipidmaps.org/tools/\)\) and the Scripps Center for Metabolomics METLIN Database \(<http://metlin.scripps.edu/>\) were also conducted in order to corroborate the assumption. The searches, although not entirely accurate, included in their many results these two fatty acids. Retention times observed for these two peaks in the different samples, eluting almost at 30 min and always after lignoceric acid \( \$C\_{24:0}\$ \) and eluting first  \$m/z\$  381 and then 395 are also keys for the probable identity of these peaks.](http://</a></p>
</div>
<div data-bbox=)

Some authors have reported the presence of very long chain fatty acids (VLFAs) in some species of the family Leguminosae. Sayed, Afifi, and Hassan (1980) reported that *Vicia sativa* subsp. *nigra*, *Vicia calcarata* and *Vicia faba* var. *minor* contain considerable amounts of higher molecular weight saturated fatty acids.



Castellón et al. (2003) also analysed the biochemical composition of the seed of six cowpea cultivars and found low percentages of pentacosanoic and arachidic acids. Other example includes groundnut oil, which according to Dean, Davis, and Sanders (2011) contains cerotic acid into its composition. No information regarding the presence of VLFA in Canola was found.

Finally the peaks with  $m/z$  values of 325 and 353 detected in pea oil samples from seeds of varieties 45760 and Galena were not determined as of lack of standards and MS/MS data.

Nevertheless, with this identification attempt some peaks were still used to localise double bonds, distinguish between positional isomers such as  $\alpha$ - and  $\gamma$ -linolenic acid and make more accurate recognitions of fatty acids than with just a single stage MS experiment.

### 3.3. RP-LC of fatty acids

The elution pattern for each fatty acid was influenced by the carbon chain length and degree of unsaturation with long chain fatty acids having higher RT than short chain fatty acids but this effect was partially offset by the effect of decreasing RT with increasing degree of unsaturation (Chen & Chuang, 2002; Seiwert et al., 2009).

The most difficult fatty acid pairs resolved were the positional isomers of linolenic acid ( $C_{18:3}$ ) with myristic acid ( $C_{14:0}$ ) followed by the behenic ( $C_{22:0}$ ) and nervonic ( $C_{24:1}$ ) acids pair. Retention times of fatty acids may be predicted by semi-empirical means (Chen & Chuang, 2002). It is often believed that the equivalent chain length (ECL) of fatty acids with  $N$  carbon atoms and  $n$  double bonds can be expressed as:

$$ECN = \text{Chain length} - (2)(\text{Number of double bonds})$$

Components with the same ECN will elute closely to each other and be considered as “critical pairs” for the separation (Purdon, 1993). According to Chen and Chuang (2002) each double bond produces a retention time approximating that of two fewer methylene groups; this therefore results in close elution of fatty acids such as  $C_{14:0}$ ,  $C_{16:1}$ ,  $C_{18:2}$  and so on. These “critical pairs” also include  $C_{16:1}$  and  $C_{20:4}$ , cis and trans  $C_{18:1}$ ,  $C_{20:0}$  and  $C_{22:1}$ , and  $C_{22:0}$  and  $C_{24:1}$ . Although there have been several attempts to separate

critical pairs of fatty acids, simultaneous separation of these molecules with different chain lengths and unsaturation in addition to cis/trans isomers has been proven difficult (Purdon, 1993).

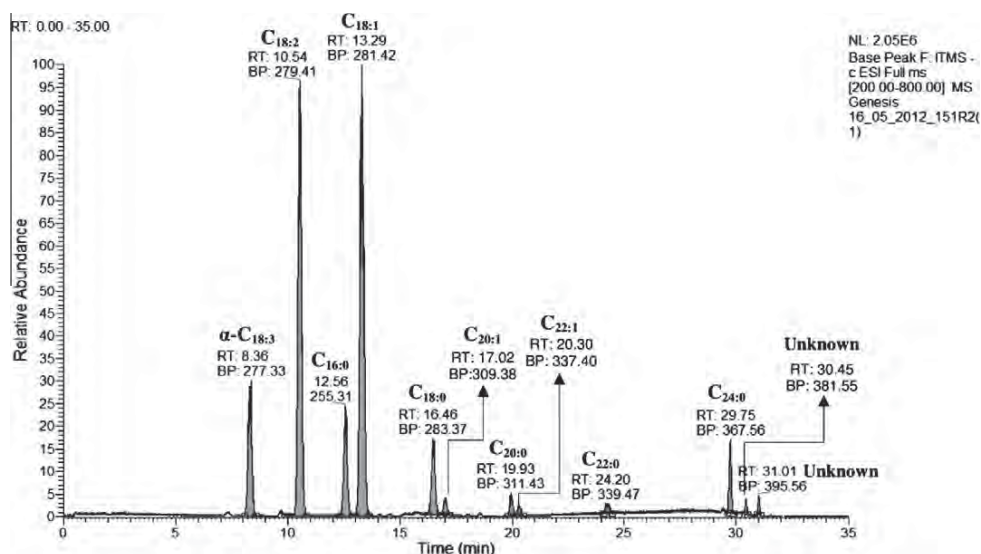
Separation based on double bond position can also be successful in limited cases. For example, oleic acid ( $C_{18:1}$ , 9c) can be separated from petroselinic acid ( $C_{18:1}$ , 6c), but is not usually separated from vaccenic acid ( $C_{18:1}$ , 11c) (Purdon, 1993). Identification of positional isomers of linolenic acid by MS has been reported as unsatisfactory and therefore chromatographic separation is necessary (Laakso, 1997). In the present study the isomers of linolenic acid tested were partially resolved between each other with  $\alpha$ -linolenic acid molecules eluting first then the corresponding  $\gamma$ -linolenic acid.

The HPLC elution strategy employed in this work was based on the studies by Li et al. (2001) and Chen and Chuang (2002). In Li et al. (2001) a gradient from 100% mobile phase A (25% acetonitrile + 75% water + 0.12% v/v acetic acid) to 100% mobile phase B (100% acetonitrile + 0.12% v/v acetic acid) linearly over 80 min was acceptable to separate the critical pairs up to  $C_{18:3}$ , however when trying to separate longer carbon chain fatty acids the elution failed (data not shown). Chen and Chuang (2002) investigated the effect of temperature and organic solvent composition on the RP separation of coumarin derivatised fatty acids according to their carbon number (from  $C_{14}$  to  $C_{22}$ ), the degree of unsaturation and their trans/cis configuration and found that methanol as organic modifier is able to separate long chain fatty acid pairs and structural isomers by altering the ECL values of the molecules. For that reason, the elution program also included methanol as a third mobile phase.

The capability of the proposed method was examined with the pea, soybean and canola samples previously prepared. Fig. 3 is a representative example of the separation of fatty acids achieved in pea oil.

Individual fatty acids were identified based on the MS/MS spectra patterns and on retention times. The isomer  $\alpha$ - $C_{18:3}$  was detected and is in general agreement with compositions of pea, soybean and canola oil that have been investigated using standard GC-FID or GC-MS techniques (Hammond et al., 2005; Shahidi, 1990, chap. 6; Yoshida et al., 2007).

Although retention times of the standards were different than those in the biological samples the same elution pattern was observed. This shift in retention can be due to changes in the pH of



**Fig. 3.** Separation of fatty acids from the oil of field pea 29595. In all screened pea samples a similar fatty acid profile was appreciated. No relevant contaminant peaks were seen. Scanned mass range was from 200.00 to 800.00. Identifications of individual fatty acids were based on MS/MS spectrum profile and retention times.

the mobile phases or due to the number of compounds present in the standard solution in opposition to the samples as peak spacing can change if more compounds of similar structure are present.

#### 4. Linearity of the response detector and limits of detection

The response of the system was established by measuring peak areas for fatty acid standards over a 5-fold range of concentration. This response was linearly dependent on the amounts of analyte up to 100 ng, with each standard calibration curve showing  $R^2$  values of more than 0.90.

Limits of detection (LODs) and quantification (LOQs) were calculated based on the standard deviation of the response (peak areas) and the slope values of each standard calibration curve and by multiplying the quotient of both by 3.3 or 9 for LODs and LOQs respectively. These data are presented in Table 3.

#### 5. Conclusions

A simple and rapid methodology by LC/ESI-MS/MS for the characterisation of fatty acids from pea oil samples was developed. The procedure was also tested with soybean and canola oils showing the possible application for oilseed investigation of fatty acids. Sample preparation and manipulation was minimal and extracts were clean enough for analysis. Compared to methods such as GC-FID or LC-UV/Vis, chemical derivatisation was not necessary thus decreasing the likelihood of unwanted effects like racemisation and/or oxidation. Generally in raw pea the following fatty acid composition was obtained: C18:2 > C18:1 > C16:0 > C18:3 > C18:0, which is consistent with literature. This fatty acid profile suggests that the species of field pea investigated may have the potential to be used as raw materials to develop a commercially accepted oilseed crop as oil composition of field pea perfectly fits for the food industry.

#### Acknowledgments

This work is being supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) (project no. 120497570), the Consortium de Recherche et Innovations en Bioprocédés Industriels au Québec (CRIBIQ), Lefsrud Seed and Processors Ltd and Agrocentre Belcan.

Special thanks to Dr. Eric Huang, McGill University, for helping with the technical and software handling of the LC-MS equipment used.

Manuel Villalobos would like to thank the Mexican National Council on Science and Technology (CONACYT) and the Fund for the Development of Human Resources (FIDERH) for financing his studies in Canada.

#### References

- Bağcı, E., & Şahin, A. (2004). Fatty acid patterns of the seed oils of some *Lathyrus* species L. (Papilionideae) from Turkey, a chemotaxonomic approach. *Pakistan Journal of Botany*, 36, 403–413.
- Balogun, A., & Fetuga, B. (1985). Fatty acid composition of seed oils of some members of the Leguminosae family. *Food chemistry*, 17(3), 175–182.
- Byrdwell, W. C. (2005). *Modern methods for lipid analysis by liquid chromatography/mass spectrometry and related techniques*. AOCS Press.
- Cahoon, E. B. (2003). Genetic enhancement of soybean oil for industrial uses: Prospects and challenges. *Journal of Agrobiotechnology Management & Economics*, 6(1–2), 11–13.
- Castellón, R. E. R., Araújo, F. M. M. C., Ramos, M. V., Andrade Neto, M., Freire-Filho, F. R., Grangeiro, T. B., et al. (2003). Biochemical composition and characterisation of lipid fraction of six cowpea cultivars. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 7(1), 149–153.
- Chen, S. H., & Chuang, Y. J. (2002). Analysis of fatty acids by column liquid chromatography. *Analytica Chimica Acta*, 465(1–2), 145–155.
- Coxon, D. T., & Wright, D. J. (1985). Analysis of pea lipid content by gas chromatographic and microgravimetric methods. Genotype variation in lipid content and fatty acid composition. *Journal of the Science of Food and Agriculture*, 36(9), 847–856.
- Dean, L. L., Davis, J. P., & Sanders, T. H. (2011). Groundnut (peanut) oil. In F. D. Gunstone (Ed.), *Vegetable oils in food technology: Composition, properties and uses* (pp. 225–242). Oxford: Blackwell Publishing Ltd.
- El-Saied, H. M., Amer, M., & Gabran, A. (1981). Unsaponifiable matter and fatty acid composition of pea oil. *Zeitschrift für Ernährungswissenschaft*, 20(2), 132–138.
- Hammond, E. G., Johnson, L. A., Su, C., Wang, T., & White, P. J. (2005). *Soybean oil*. Bailey's Industrial Oil and Fat Products.
- Hilditch, T. P., & Williams, P. N. (1964). *The chemical constituents of natural fats* (4th ed.). London: Chapman and Hall.
- Holman, R. T. (1981). Essential fatty acids in nutrition and disease. *Chemistry & Industry*, 21, 704–709.
- Kerwin, J. L., Wiens, A. M., & Ericsson, L. H. (1996). Identification of fatty acids by electrospray mass spectrometry and tandem mass spectrometry. *Journal of mass spectrometry*, 31(2), 184–192.
- Kurata, S., Yamaguchi, K., & Nagai, M. (2005). Rapid discrimination of fatty acid composition in fats and oils by electrospray ionization mass spectrometry. *Analytical sciences*, 21(12), 1457–1465.
- Laakso, P. (1997). Characterisation of  $\alpha$ - and  $\gamma$ -linolenic acid oils by reversed-phase high-performance liquid chromatography-atmospheric pressure chemical ionization mass spectrometry. *Journal of the American Oil Chemists' Society*, 74(10), 1291–1300.
- Li, Z., Gu, T., Kelder, B., & Kopchick, J. (2001). Analysis of fatty acids in mouse cells using reversed-phase high-performance liquid chromatography. *Chromatographia*, 54(7–8), 463–467.
- Lima, E., & Abdalla, D. (2002). High-performance liquid chromatography of fatty acids in biological samples. *Analytica Chimica Acta*, 465(1), 81–91.
- Murcia, M., & Rincon, F. (1992). Size as source of variance in lipid composition of pea. *Food chemistry*, 44(1), 29–35.
- Nikolopoulou, D., Grigorakis, K., Stasini, M., Alexis, M., & Iliadis, K. (2007). Differences in chemical composition of field pea (*Pisum sativum*) cultivars: Effects of cultivation area and year. *Food chemistry*, 103(3), 847–852.
- Nishiyama-Naruke, A., Souza, J., Carneiros, M., & Curí, R. (1998). HPLC determination of underivatized fatty acids saponified at 37 °C analysis of fatty acids in oils and tissues. *Analytical letters*, 31(14), 2565–2576.
- Perret, D., Gentili, A., Marchese, S., Sergi, M., & Caporossi, L. (2004). Determination of free fatty acids in chocolate by liquid chromatography with tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, 18(17), 1989–1994.
- Pulse Canada. Canadian pulse industry: Situation and outlook. (2012). URL: <http://www.agr.gc.ca/pol/maddam/index\_e.php?s1=pubs&s2=spec&s3=php&page=spec\_2012-06-14> Accessed 30.07.12.
- Purdon, M. (1993). Application of HPLC to lipid separation and analysis: Sample preparation. In E. G. Perkins (Ed.), *Analyses of fats, oils and derivatives* (pp. 145). AOCS.
- Ratnayake, W., Hoover, R., Shahidi, F., Perera, C., & Jane, J. (2001). Composition, molecular structure, and physicochemical properties of starches from four field pea (*Pisum sativum* L.) cultivars. *Food chemistry*, 74(2), 189–202.
- Ryan, E., Galvin, K., O'Connor, T., Maguire, A., & O'Brien, N. (2007). Phytosterol, squalene, tocopherol content and fatty acid profile of selected seeds, grains, and legumes. *Plant Foods for Human Nutrition (Formerly Qualitas Plantarum)*, 62(3), 85–91.
- Sayed, D., Afifi, M. S., & Hassan, M. A. (1980). A study of lipid content of the leaves, stems and seeds of *Vicia sativa* L., *Vicia calcarata* Desf. and *Vicia faba* L. (var. minor) growing in Egypt. *Egyptian Journal of Pharmaceutical Sciences*, 21(12), 43–52.
- Seiwert, B., Givalisco, P., & Willmitzer, L. (2009). Advanced mass spectrometry methods for analysis of lipids from photosynthetic organisms. In H. Wada & N. Murata (Eds.), *Lipids in photosynthesis* (pp. 445–461). Springer.
- Shahidi, F. (1990). *Canola and rapeseed: Production, chemistry, nutrition, and processing technology* (1st ed.). New York: Van Nostrand Reinhold.
- Singh, N., Kaur, N., Rana, J. C., & Sharma, S. K. (2010). Diversity in seed and flour properties in field pea (*Pisum sativum*) germplasm. *Food chemistry*, 122(3), 518–525.
- Thiessen, D. L. (2004). Optimisation of feed peas, canola and flaxseed for aqua feeds: The Canadian prairie perspective. *Avances en Nutricion Acuicola VII. Memorias del VII Simposium Internacional de Nutricion Acuicola* (pp. 16–19).
- Wang, N., & Daun, J. K. (2004). *The Chemical composition and nutritive value of Canadian pulses* (pp. 19–29). Canadian Grain Commission (CGC).
- Welch, R. W., & Wynne Griffiths, D. (1984). Variation in the oil content and fatty acid composition of field beans (*Vicia faba*) and peas (*Pisum spp.*). *Journal of the Science of Food and Agriculture*, 35(12), 1282–1289.
- Yoshida, H., Tomiyama, Y., Tanaka, M., & Mizushima, Y. (2007). Characteristic profiles of lipid classes, fatty acids and triacylglycerol molecular species of peas (*Pisum sativum* L.). *European Journal of Lipid Science and Technology*, 109(6), 600–607.

## ADVANCES AND CHALLENGES WITH MICRO-IRRIGATION<sup>†</sup>

CHANDRA A. MADRAMOOTOO AND JANE MORRISON\*

*Brace Centre for Water Resources Management, McGill University, Macdonald Campus, Sainte-Anne-de-Bellevue, Quebec, Canada*

### ABSTRACT

As global concerns surrounding water scarcity and food security escalate, there will be more demand for micro-irrigation to meet growing food demands. Micro-irrigation offers many advantages over conventional irrigation methods, including the ability to apply limited amounts of water directly to the crop root zone, incorporation of fertigation, reduced weed and pest infestation, and lower capital and operating costs. In recent decades, there has been considerable growth in the acreage under micro-irrigation, mainly as a result of lower costs, improvements in filtration and emitter technology, and increased grower confidence in the technology. Research advances and technological improvements have made micro-irrigation applicable to a more diverse set of applications, cropping systems, and water quality conditions. Cost and availability of water are also major drivers. Research in nano- and biofiltration techniques, soil moisture sensors, and precision irrigation shows great promise for the advancement of micro-irrigation. Nevertheless, several technological challenges remain, especially for non-row or non-orchard crops, and in regions where water quality is severely impaired. Innovations in these areas are required, as well as a transfer of the technology to small farmers in water-scarce regions who traditionally surface irrigate. Copyright © 2013 John Wiley & Sons, Ltd.

KEY WORDS: micro-irrigation; drip irrigation; water use efficiency; agriculture; food security

*Received 16 July 2012; Revised 3 October 2012; Accepted 3 October 2012*

### RÉSUMÉ

En réponse aux préoccupations grandissantes concernant la rareté de l'eau et la sécurité alimentaire et aux enjeux afférents, on peut s'attendre à une augmentation de la demande de micro-irrigation pour faire face aux besoins alimentaires croissants. La micro-irrigation offre plusieurs avantages lorsqu'elle est comparée aux méthodes conventionnelles d'irrigation. Cela inclut la capacité d'appliquer des quantités limitées d'eau directement dans la couche arable des cultures, l'incorporation de la fertirrigation, la réduction de l'infestation par des espèces nuisibles, végétales ou animales, ainsi que la réduction des coûts d'investissements et de fonctionnement. Durant les dernières décennies, il y a eu une croissance considérable des zones micro-irriguées, principalement grâce aux réductions de coûts, aux améliorations dans les technologies de filtration et d'émetteurs électroniques, et à la confiance grandissante des producteurs envers cette technologie. Les avancées de la recherche et l'amélioration des technologies ont rendu la micro-irrigation applicable à un groupe plus diversifié d'applications, de systèmes de récolte, et de conditions de qualité d'eau. Le coût et la disponibilité de l'eau sont aussi des facteurs majeurs. La recherche sur les techniques de nanofiltration et de bio-filtration, sur les capteurs d'humidité du sol, et sur l'agriculture de précision permet d'entrevoir des avancées certaines de la micro-irrigation. Cependant, il reste beaucoup de défis technologiques, surtout pour les cultures qui ne sont pas en rangées et celles qui ne font pas partie d'un verger, ainsi que dans les régions où la qualité de l'eau est gravement détériorée. Des innovations dans ces domaines sont nécessaires, ainsi qu'un transfert de technologie envers les petits cultivateurs qui utilisent traditionnellement l'irrigation de surface dans des régions pauvres en eau. Copyright © 2013 John Wiley & Sons, Ltd.

MOTS CLÉS: micro-irrigation; irrigation au goutte-à-goutte; efficacité de l'utilisation de l'eau; agriculture; sécurité alimentaire

### INTRODUCTION

The irrigation sector is under pressure to increase its efficiency since it is the major user of fresh water globally. This is exacerbated as water resources become scarcer due to climate change, and as the competition for water from other economic

\* Correspondence to: Jane Morrison, Brace Centre for Water Resources Management, McGill University, Macdonald Campus, Sainte-Anne-de-Bellevue, Quebec, Canada. E-mail: jane.morrison@mail.mcgill.ca

<sup>†</sup>Avancées et défis de la micro-irrigation.

and environmental uses increases. Demand for more efficient water applications is therefore on the rise. Micro-irrigation, which is the precise application of water on or below the soil surface at low pressure using small devices that spray, mist, sprinkle, or drip water, is becoming more attractive (Hla and Scherer, 2003). In the past, micro-irrigation required high capital investment and was therefore mostly used for high-value cash crops, such as vegetables and fruit trees, and was infrequently used for lower-value staple crops (Brouwer *et al.*, 1988). In recent decades, increased sales and improvements in technology have resulted in greater adoption of micro-irrigation worldwide. Drip irrigation is a common form of micro-irrigation; Table I shows the worldwide growth in drip irrigation since its conception in the 1970s. Table II displays the rapid growth in acreage under drip irrigation within selected countries between the years 1974 and 2006.

Table III compares the areas of sprinkler irrigation and micro-irrigation by country, and their percentage of the total irrigated area.

In some countries, such as Spain, Korea, Israel, Poland, and Chile, the area cropped to drip irrigation is higher than that which is sprinkler irrigated (Table III). It is also worthy of note that India's percentage of drip irrigated area is increasing and is not far behind its sprinkler irrigated area (ICID, 2011). Several government and non-government organizations are actively promoting micro-irrigation in developing countries (Varma *et al.*, 2006).

The high-value horticultural, ornamental, and landscape industries have taken advantage of micro-irrigation, and remain its biggest users. Small-scale cash crop growers have

expanded their production by adopting drip irrigation. This is particularly evident in the case of vegetable production in periurban areas in some parts of Africa and Asia. However, more innovations are required to expand the use of drip irrigation for non-orchard crops such as rice, wheat, and maize. In order to significantly expand the area under micro-irrigation, these technological innovations will have to come with lower capital costs and the ability to irrigate with poorer quality water.

## ADVANTAGES AND BENEFITS

Micro-irrigation provides a constant supply of water in the crop root zone and has been proven to provide a higher crop yield and increased water use efficiency over conventional irrigation methods. Micro-irrigation systems offer a high level of control over water applications. These systems are also very economical as they use a low volume of water, at low pressure, resulting in low energy costs (Varma *et al.*, 2006). With micro-irrigation there is no percolation to groundwater and there is also no surface runoff. In addition, because of a reduced wetted area, less water is lost to evaporation.

Due to its high water use efficiency, micro-irrigation is increasingly being used as a strategy to address water scarcity and poverty. Researchers, such as Shah (2011), have pointed to the water savings and yield increases due to micro-irrigation (Table IV). In addition to farm productivity (crop yield and output), farmer income and food security are also increased. With earlier harvests, labour costs are reduced. Improvements in drip irrigated crop quality have also been observed (Madramootoo and Rigby, 1991).

Generally, micro-irrigation systems are easy to install and manage, and can be used for a variety of orchard and row crops, particularly for fruit and vegetables. Land levelling is not necessary because system components, pressure regulating valves and pressure compensating emitters, allow for a uniform water distribution and application rate. A major innovation has been the development of fertilizer injectors which can be attached to micro-irrigation systems to allow for fertigation, thereby improving crop nutrient management. Both water and fertilizer can be applied throughout the growing season in judicious amounts to match crop requirements. Additionally, incidence of pest and weed invasion, and other plant diseases, occurs less frequently due to a reduced wetted area and drier soil surface. By reducing the labour needed to protect crops, and by reducing pesticide/herbicide use, there is an overall financial saving to the farmer (Varma *et al.*, 2006).

## MAJOR ADVANCES

Major advances have taken place in micro-irrigation technology, resulting in increased adoption and a growing number of

Table I. Worldwide growth in drip irrigation (Postel *et al.*, 2001)

Year	Global acreage under drip irrigation (million ha)
1970s	0.056
1991	1.6
2001	≈2.8

Table II. Acreage under drip irrigation for selected countries (ha)

Country	Acreage under drip irrigation		
	1974–1981	1991	2000–2006
China	n.a.	19 000 <sup>a</sup>	759 500 <sup>b</sup> (2006)
India	20 <sup>b</sup> (1981)	70 680 <sup>b</sup>	578 200 <sup>b</sup> (2004)
Israel	6 073 <sup>c</sup> (1974)	104 300 <sup>b</sup>	168 800 <sup>b</sup> (2004)
South Africa	3 482 <sup>c</sup> (1974)	102 300 <sup>a</sup>	178 000 <sup>b</sup> (2000)
United States	29 150 <sup>c</sup> (1974)	606 000 <sup>a</sup>	1 615 000 <sup>b</sup> (2005)

<sup>a</sup>Postel (1992).

<sup>b</sup>Food and Agriculture Organization of the United Nations (2012).

<sup>c</sup>Gustafson (1974).



Table III. Areas of sprinkler and micro-irrigation for selected countries (ICID, 2011<sup>†</sup>)

Country	Total irrigated area (million ha)	Sprinkler irrigation		Micro-irrigation		Year of reporting
		Area (ha)	% of total irrigated area	Area (ha)	% of total irrigated area	
USA	24.70	12 300 000	50.0	1 640 000	6.6	2009
India	60.90	3 040 000	5.0	1 900 000	3.1	2010
China	59.30	2 930 000	4.9	1 670 000	2.8	2009
Russia	4.50	3 500 000	77.8	20 000	0.4	2008
Brazil	4.45	2 410 000	54.2	328 000	7.4	2006
Spain	3.41	733 000	21.5	1 630 000	47.8	2010
Italy	2.67	981 000	36.8	571 000	21.4	2010
France	2.90	1 380 000	47.6	103 000	3.6	2011
South Africa	1.67	920 000	55.1	365 000	21.9	2007
Saudi Arabia	1.62	716 000	44.2	198 000	12.2	2004
Iran	8.70	460 000	5.3	270 000	3.1	2009
Australia	2.55	524 000	20.6	191 000	7.5	2000
Canada	0.87	683 000	78.5	6 030	0.7	2004
Mexico	6.20	400 000	6.5	200 000	3.2	1999
Korea	1.01	200 000	19.8	400 000	39.6	2009
Israel	0.23	60 000	26.0	170 000	73.6	2000
Morocco	1.65	190 000	11.5	8 250	0.5	2003
Syria	1.28	93 000	7.3	62 000	4.8	2000
Turkey	5.34	110 000	2.1	26 000	0.5	2009
United Kingdom	0.11	105 000	95.5	6 000	5.5	2005
Portugal	0.63	40 000	6.4	25 000	4.0	1999
Chile	1.09	16 000	1.5	23 000	2.1	2006
Chinese Taipei	0.38	18 900	5.0	8 750	2.3	2009
Bulgaria	0.59	21 000	3.6	3 000	0.5	2008
Czech Rep.	0.15	11 000	7.2	5 000	3.3	2007
Philippines	1.52	7 180	0.5	6 640	0.4	2004
Poland	0.10	5 000	5.0	8 000	8.0	2008
Malaysia	0.38	2 000	0.5	5 000	1.3	2009
Estonia	0.001	500	50.0	500	50.0	2010

advantages. In addition, innovations allow for a more diverse set of applications.

Clogging and inefficient water filtration have traditionally been major issues when employing micro-irrigation. The invention of functional and cost-effective sand and screen filtration devices has helped to overcome this issue, and has ultimately resulted in an expanded use of micro-irrigation in areas with low-quality water.

Micro-irrigation can be combined with precision irrigation scheduling, enabling the application of water at a time and rate that are based on a crop's precise water requirement. With precision irrigation, scheduling is synchronized with weather and soil conditions, and a crop's evapotranspiration rate (Jones, 2004). Precision irrigation saves water and money, and reduces runoff and energy consumption. In addition, by irrigating crops with a precise amount of water suited to the crop's growth stage, there is also the potential to increase crop yield (Cooley *et al.*, 2009).

The development of time domain and frequency domain reflectometry sensors which provide information on soil moisture levels on a real-time basis has enabled the automation of

irrigation systems. In the past, farmers decided when to irrigate based on intuition, weather, and physical plant/soil conditions. Sensors used for real-time irrigation scheduling take into account crop type, growth stage, soil type, and soil moisture status in order to take the guesswork out of irrigating and provide farmers with a more efficient and accurate way of determining when to apply water and for how long. Modern equipment allows for real-time data on soil water status to be available directly in the field, enabling farmers to make timely irrigation decisions. In addition, with advanced technology, irrigation operations can be automatically triggered, consistently applying an accurate amount of water to crops (avoiding unnecessary excess or deficits in water applications). Recent work by Jaria and Madramootoo (2011) has shown how automated soil moisture sensors and weather measurements can be used in a computerized system, to determine irrigation triggers, thereby permitting the farmer to better and more precisely schedule drip irrigation applications.

Improved knowledge about the interactions between the plant, soil, and water continuum has paved the way for innovative water regulating mechanisms. For example, with



Table IV. Drip vs surface irrigation—water saving and increase in yield (Shah, 2011)

Crop	Yield (kg ha <sup>-1</sup> )			Irrigation		
	Surface	Drip	Increase (%)	Surface	Drip	Saving (%)
Beet root	570	880	54	86	18	79
Bitter gourd	3 200	4 300	34	76	33	57
Brinjal	9 100	14 800	63	168	64	62
Broccoli	14 000	19 500	39	70	60	14
Chili	17 100	27 400	60	27	18	33
Cucumber	4 230	6 090	44	109	42	62
Okra	15 500	22 500	45	54	24	56
Onion	28 400	34 200	20	52	26	50
Potato	17 200	29 100	69	60	28	54
Radish	1 050	1 190	13	46	11	76
Sweet potato	4 240	5 890	39	63	25	60
Tomato	6 180	8 870	44	50	11	79
Banana	57 500	87 500	52	176	97	45
Grapes	26 400	32 500	23	53	28	47
Papaya	13 000	23 000	77	228	73	68
Pomegranate	3 400	6 700	97	21	16	24
Watermelon	8 210	50 400	514	72	25	65

deficit irrigation, water is applied below the crop's full water requirement (a percentage of a crop's evapotranspiration rate), reducing irrigated water use and conserving water. Water use effectiveness is increased by stressing crops, but only to a point which does not adversely affect yields (Feres and Soriano, 2007).

Partial root zone drying is another strategy that can be employed with micro-irrigation for a more efficient use of irrigation water. The process requires that half of a crop's rooting system be exposed to alternate wetting and drying; essentially half of the root system is put in a drying state while the remainder remains irrigated. The alternating frequency is a function of the crop type, growth stage, and specific water requirement. This technique is based on an understanding of the mechanisms controlling transpiration, and requires sophisticated management and highly accurate monitoring (McCarthy *et al.*, 2002). With this process there is a potential to reduce water use while increasing canopy vigour and maintaining yields in regions of water scarcity.

Micro-irrigation technology is enabling new and diverse methods of providing dependable irrigation in water-scarce regions. Treated and untreated wastewater can be applied in a manner which targets only suitable crops. Micro-irrigation is more appropriate than other irrigation methods for the reuse of wastewater because there is no aerosol generation and no wastewater comes into contact with plant foliage. There are also fewer problems with odours, ponding, and runoff. In addition, studies suggest that the nitrogen present in wastewater is better absorbed by plants and less likely to

pollute groundwater when applied directly to plant roots. When using wastewater with micro-irrigation it is necessary to ensure that emitters do not clog. Systems must be closely monitored to ensure a uniform application and full functionality (Casey *et al.*, 1999).

Another strategy for overcoming water shortages is the use of reverse osmosis in subsurface drip irrigation lines. Reverse osmosis technology enables the use of brackish water, frequently found in groundwater aquifers, as a source of irrigation water; allowing water that would normally be deemed unsuitable to be used for irrigating food crops. Desalination is achieved using a semi-permeable reverse osmosis membrane within the drip line, supplying purified water to irrigated crops and preventing the accumulation of salts in the soil. Unlike conventional reverse osmosis systems which require the use of high-energy pumps, the transport of water across the reverse osmosis membrane within a drip line is driven by the matrix potential of the soil. Reverse osmosis subsurface drip irrigation systems are low flow, low pressure, and have an especially high water use efficiency because water extraction is a direct function of the water requirement of the crop (Leslie, 2010).

## FUTURE CHALLENGES

There remain several obstacles which must be overcome in order to increase the adoption of micro-irrigation, particularly in developing countries where small farms predominate.

Micro-irrigation is best suited for irrigating individual plants or trees, or row crops, but is generally deemed unsuitable for close-growing crops such as rice or other cereals (Brouwer *et al.*, 1988). As a result, farmers in regions which rely on these types of close-growing crops have been unable to benefit from micro-irrigation. However, with advances in technology, micro-irrigation is being adapted for a wider variety of crops. The driver is to increase water use efficiency and food security in water-scarce regions and developing countries which depend on cereal crops as dietary staples.

Rice, which is traditionally flood irrigated, uses large volumes of water. With drip irrigation the water requirement is 50–60% less. Drip irrigation keeps the moisture availability in the soil continuously close to the crop's water requirement. However, non-flooding leads to increased weed presence, making herbicide use a critical element when using drip irrigation in rice farming (Soman, 2012). More knowledge about the specific water requirements of different cereal crops and more adapted equipment and technology are needed in regions which rely on these types of crops.

The benefits of using a buried drip irrigation system as opposed to a surface drip irrigation layout are being examined. On one hand, buried systems have a longer lifespan as they are less likely to be damaged or tampered with by fieldworkers and rodents. Studies have shown that with buried drip lines, crop yields are equal to, or better than, those in which surface drip lines are installed. In addition, the water and fertilizer requirements for buried systems are equal to or lower than those for surface irrigation. However, some subsurface drip systems are complex, requiring air entry ports and flushing manifolds. The ability for subsurface irrigation to maintain a dry soil surface is an advantage for some crop situations (i.e. to reduce weed growth), while it is a disadvantage in other situations (i.e. for germination of shallow-planted seeds). An additional benefit is the use of subsurface irrigation in wastewater applications. This results in fewer odour problems, and the deeper placement of phosphorus in the soil profile facilitates plant uptake (Camp *et al.*, 2000).

As we move to automated irrigation scheduling, more research is required to determine the most appropriate placement of soil moisture sensors. Currently, sensors are only able to provide soil water status at a particular point and depth at a given site. As such, placement of the sensor is important in order to provide a representative measurement for making comprehensive irrigation decisions. Placement of the sensor in terms of the wetted area in relation to the crop root zone is crucial for triggering irrigation systems on and off.

An evolving area of research is the application of satellite-positioning systems, geographic information systems, automated machine guidance, and remote-sensing technologies for precision irrigation scheduling and improved agricultural water management (Joseph and Morrison, 2006). Thermal imagery using low-flying unmanned aerial vehicles, drones,

aircraft, or satellites can determine the water content of soil or plant water status at high spatial resolution, and consequently trigger irrigation precisely in a specific location. As a result, water conservation and crop productivity can be optimized. This technology can overcome obstacles associated with soil heterogeneity and field variability in order to ensure individual areas are being irrigated according to their specific water requirements. However, the technology is still being refined in order to overcome issues associated with cloud interference or the need for frequent imaging. Systems which utilize crop simulation models in addition to remote sensing have been proposed in order to overcome these obstacles and effectively manage precision or variable rate irrigation (Barnes *et al.*, 2000). Further research is required to improve the accuracy and cost-effectiveness of these techniques.

There is potential to capitalize on the use of nanotechnology and biotechnology in micro-irrigation, particularly with respect to water quality improvement, filtration techniques, and reduced emitter clogging. Nanomaterial-based biosensors can recognize, measure, and monitor the presence of contaminants, with the potential to operate on site, in real time. Nanofiltration membrane technology enables the identification and trapping of particles at the nanoscale, and the capability for partial desalination. In addition, there is reduced fouling of membranes and filters. Nanotechnology is also used in soil moisture sensor design, achieving a high level of accuracy, rapid response rates, compact size, and robustness. With these innovations, the challenge lies in transferring research into practice. Improved cost-effectiveness, availability, and informed management will help to increase the adoption of these promising technologies (Organisation for Economic Co-operation and Development (OECD), 2011).

Although technological development and simplification have reduced the overall cost of micro-irrigation systems, the technology is still expensive. Additionally, a high level of technical skill is required for proper system design, maintenance, and optimal efficiency. Advancements to further reduce costs, labour, and to facilitate the initial system design and installation will make this technology increasingly attractive to farmers.

Due to the high initial investment required for taking up or switching to micro-irrigation systems, small farmers in developing countries have been slow to adopt this practice. In addition, a lack of cash and expertise, and the crop specificity associated with available micro-irrigation technologies, further reduce its desirability (Varma *et al.*, 2006). Studies have shown that farmers are rarely motivated by the water and energy savings potential of new technologies. In many countries, water is not metered and power is subsidized, giving farmers no financial incentive to strive for resource conservation. Furthermore, in countries where farmers have practised surface irrigation for decades and are accustomed to having water delivered in canals, they see no immediate benefit to culturally changing

to other methods of irrigation. Farmers tend to look for new technologies which promise increased profit (Soman, 2012). As such, strong institutional support, promotion, product subsidies, and improved access to credit are needed for an increased uptake of micro-irrigation technology in developing countries (Varma *et al.*, 2006).

## CONCLUSIONS

Since the introduction of micro-irrigation technology over 40 years ago, there is now close to some 3 million ha of drip irrigation worldwide. Micro-irrigation has been particularly successful for horticultural, ornamental, and landscape applications, and has been applied to a wide range of climatic conditions from humid to arid and semi-arid regions. Its advantages with respect to water and energy savings, reduced labour, and improved fertilizer application through fertigation are well recognized. Micro-irrigation will continue to attract attention as water resources become increasingly limited. Its suitability for sloping and undulating topography, and applicability to any size of farm, offer several advantages compared to other irrigation techniques. Advances in emitter and dripper technology, the introduction of inexpensive drip tape, and the development of low-cost sand and screen filters have also helped to expand the acreage under micro-irrigation. Additionally, new soil moisture sensors have enabled the automation of these irrigation systems.

Ongoing research with bio- and nanofiltration systems, particularly the incorporation of nanofilters in the emitters and drip lines, will allow micro-irrigation to be more widely used in regions that suffer from poor water quality, and in situations where wastewater is used for irrigation. The latter will become increasingly important in regions affected by water scarcity. The use of soil moisture sensor technology, together with remote sensing of soil and crop parameters, geographic information systems, global positioning systems, and crop growth models, will allow for improved water management and water use efficiency under micro-irrigation. The goal is to make precision irrigation or variable rate irrigation an essential facet of micro-irrigation.

A significant challenge is to apply drip irrigation technology to the production of cereal crops, particularly in developing countries. In these parts of the world, there are several social, technical, and institutional challenges which must be overcome. Education and knowledge transfer must be accelerated. Energy and water subsidies must be removed to see the full benefits of drip irrigation.

## ACKNOWLEDGEMENTS

This article is based on a keynote presentation given by the senior author at the 8th Micro-irrigation Congress, held in

Tehran on 21 October 2011. The senior author is grateful to the Congress organizers for this invitation. The senior author was President of ICID at the time of the Congress. Special thanks go to Mr Felix Jaria, graduate student at the Brace Centre for Water Resources Management, McGill University, who helped to compile some of the statistics for the paper.

The senior author gratefully acknowledges the following agencies which generously support his research programme in irrigation and water management: McGill University, Canadian Foundation for Innovation, Natural Sciences and Engineering Research Council of Canada, Canadian International Development Agency, International Development Research Centre of Canada, Agriculture and Agri-Food Canada, Max Bell Foundation, Fonds de recherche du Québec-Nature et technologies, Ontario Ministry of Agriculture, Food and Rural Affairs, and the Canadian Water Network.

## REFERENCES

- Barnes EM, Pinter PJ Jr, Kimball BA, Hunsaker DJ, Wall GW, LaMorte RL. 2000. Precision irrigation management using modeling and remote sensing approaches. Presented at the 4th Decennial National Irrigation Symposium. American Society of Agricultural Engineers, Phoenix, Arizona, 14–16 November.
- Brouwer C, Prins K, Kay M, Heibloem M. 1988. *Irrigation Water Management: Irrigation Methods*. Food and Agriculture Organization of the United Nations: New York, USA.
- Camp CR, Lamm FR, Evans RG, Phene CJ. 2000. Subsurface drip irrigation—past, present, and future. Presented at the 4th Decennial National Irrigation Symposium. American Society of Agricultural Engineers, Phoenix, Arizona, 14–16 November.
- Casey P, Lake A, Falvey C, Ross JA, Frame K. 1999. *Spray and Drip Irrigation for Wastewater Reuse, Disposal*. National Small Flows Clearing House, West Virginia University: Morgantown, W.Va.
- Cooley H, Christian-Smith J, Gleick PH. 2009. *Sustaining Agriculture in an Uncertain Future: the Role of Water Efficiency*. Pacific Institute: Oakland, Calif., USA.
- Fereres E, Soriano MA. 2007. Deficit irrigation for reducing agricultural water use. *Journal of Experimental Botany* **58**(2): 147–159.
- Food and Agriculture Organization of the United Nations. 2012. AQUASTAT. From AQUASTAT country database: <http://www.fao.org/nr/water/aquastat/main/index.stm>
- Gustafson CD. 1974. Drip irrigation in the USA. Presented at the Sprinkler Irrigation Association Annual Technical Conference, Denver, Colo., USA.
- Hla AK, Scherer TF. 2003. Introduction to Micro-Irrigation. North Dakota State University Extension Service, AE-1243: Fargo, N.D., USA.
- International Commission on Irrigation and Drainage (ICID). 2011. Sprinkler and micro-irrigated areas in some participating members of ICID. From ICID database: <http://www.icid.org/database.html>
- Jaria F, Madramootoo CA. 2011. Irrigation scheduling of field tomatoes based on stand-alone continuous, real-time soil moisture sensor data. Presented at the northeast Agricultural and Biological Conference, South Burlington, Vermont, 24–27 July.
- Jones HG. 2004. Irrigation scheduling: advantages and pitfalls of plant-based methods. *Journal of Experimental Botany* **55**(407): 2427–2436.
- Joseph T, Morrison M. 2006. *Nanotechnology in Agriculture and Food*. The Institute of Nanotechnology: Glasgow, Scotland.

- Leslie G. 2010. Reverse osmosis pipe for drip irrigation. In<sup>†</sup> *ReWater: Water Recycling in Australia*. Atura Pty Ltd.: Melbourne, Australia.
- Madramootoo CA, Rigby M. 1991. Effects of trickle irrigation on the growth and sunscald of bell peppers (*Capsicum annuum* L.) in southern Quebec. *Agricultural Water Management* **19**(2): 181–189.
- McCarthy MG, Loveys BR, Dry PR, Stoll M. 2002. *Regulated Deficit Irrigation and Partial Rootzone Drying as Irrigation Management Techniques for Grapevines*. Food and Agriculture Organization of the United Nations: New York, USA.
- Organisation for Economic Co-operation and Development (OECD). 2011. *Fostering Nanotechnology to Address Global Challenges: Water*. OECD: Paris, France.
- Postel S. 1992. *Last Oasis: Facing Water Scarcity*. W. W. Norton and Company: New York.
- Postel S, Polak P, Gonzales F, Keller J. 2001. Drip irrigation for small farmers. *Water International* **26**(1): 3–13.
- Shah SK. 2011. *Towards Adopting Nanotechnology in Irrigation: Micro Irrigation Systems*. India Water Portal: Karnataka, India.
- Soman P. 2012. Drip irrigation and fertigation technology for rice cultivation. Presented at the Asian Irrigation Forum, Manila, Philippines, 11–13 April.
- Varma S, Verma S, Namara R. 2006. Promoting micro-irrigation technologies that reduce poverty. International Water Management Institute, Water Policy Briefing 23: Colombo, Sri Lanka.

# Application of the Standardized Precipitation Index and Normalized Difference Vegetation Index for Evaluation of Irrigation Demands at Three Sites in Jamaica

Johanna Richards<sup>1</sup>; Chandra A. Madramootoo<sup>2</sup>; Manish Kumar Goyal<sup>3</sup>; and Adrian Trotman<sup>4</sup>

**Abstract:** Agricultural production is a significant contributor to the economy of Jamaica, which is situated in the northwestern Caribbean Sea; this production is heavily dependent on seasonal rainfall because only approximately 10% of Jamaica's cultivated lands are irrigated. Drought is a disastrous natural phenomenon that has a significant impact on socioeconomics, agriculture, and the environment. In the 2000–2001 drought experienced in Jamaica, there were crop losses amounting up to US\$6 million. Hence, drought index information is essential for better planning for drought impacts and allows for the introduction of mitigation measures in the agricultural sector. Therefore, the objective of this paper is to evaluate the suitability of both the Standardized Precipitation Index (SPI) and the Normalized Difference Vegetation Index (NDVI) in reflecting water stressed conditions and irrigation demand requirements for three agricultural sites: Savanna-la-Mar in the parish of Westmoreland, Beckford Kraal in the parish of Clarendon, and Serge Island in the parish of St. Thomas, all in Jamaica. These sites were selected based on soil characteristics, historical rainfall data, and farming practices. The results indicate that the NDVI provides a suitable representation of these areas for only the driest months of the year, and that either the one-month or three-month SPI was found to be more representative of soil moisture conditions. Furthermore, a correlation analysis was also conducted between the SPI and soil moisture for El Niño years only, because the El Niño/Southern Oscillation phenomenon has been responsible for many of the droughts Jamaica has experienced. In these years, good correlations between soil moisture and the one-month and three-month SPI were obtained in some wet months, in addition to the dry months. This paper provides soil moisture values for all of the different categories and values of the SPI relating to water scarcity. It also provides irrigation requirements for the “moderately dry” and “severely dry” SPI drought categories. DOI: [10.1061/\(ASCE\)IR.1943-4774.0000629](https://doi.org/10.1061/(ASCE)IR.1943-4774.0000629). © 2013 American Society of Civil Engineers.

**CE Database subject headings:** Correlation; Precipitation; Vegetation; Irrigation; Caribbean.

**Author keywords:** Standardized Precipitation Index; Normalized Difference Vegetation Index; Correlation analysis; Regression; Irrigation management.

## Introduction

Drought is a slowly developing phenomenon, and although several definitions formally exist, it is typically viewed as abnormally low water availability as a result of abnormally low levels of rainfall. Drought in the West Indies is typically related to disruptions in the seasonal rainfall cycle, primarily caused by El Niño/Southern Oscillation (ENSO) (Chen et al. 2005). When an El Niño event

occurs, the climate of the West Indies is characterized by drier than normal conditions during the later months of the rainfall season. Many island-wide meteorological droughts occurring in 1965, 1969, 1972, 1976, 1982–1983, 1991, and 1997 were caused by El Niño events (Chen et al. 2005). However, in 1976 and 1991, the island received 72 and 73%, respectively, of normal total annual rainfall with respect to a 30-year mean, resulting in the worst drought conditions during that period (Chen et al. 2005).

Jamaica is an island situated in the northwestern Caribbean Sea, which is centered along latitude 18°15' N and longitude 77°20' W. The 100-year mean annual rainfall for the island (1890–1990) is 1,895 mm (Meteorological Service of Jamaica 2009). The rainfall pattern throughout most of the island is bimodal, meaning that there are two wet periods throughout the year. Agriculture in Jamaica is heavily dependent on seasonal rainfall, with which approximately 10% of its cultivated lands are irrigated. The Jamaican agricultural sector employs approximately 20% of the labor force [Food and Agriculture Organization of the United Nations (FAO) 2003; Planning Institute of Jamaica (PIOJ) 2008; Richards et al. (2013)]. Estimates of agricultural production in Jamaica indicate that 95% of this production is rain-fed (Chen et al. 2005).

Interestingly enough, La Niña events in the cold equatorial Pacific waters can induce drought conditions in the early rainfall season of the following year. This scenario has been deemed the cause of the island-wide meteorological droughts that occurred in Jamaica in 1971, 1974, 1975, 1985, 1989, and 2000 (Chen et al. 2005). During the period from December 1996 to December 1998,

<sup>1</sup>Graduate Student, Dept. of Bioresource Engineering, McGill Univ., 21,111 Lakeshore Rd., Ste. Anne de Bellevue, QC, Canada, H9X 3V9. E-mail: johanna.richards@mail.mcgill.ca

<sup>2</sup>Professor and Dean, Dept. of Bioresource Engineering, McGill Univ., 21,111 Lakeshore Rd., Ste. Anne de Bellevue, QC, Canada, H9X 3V9. E-mail: chandra.madramootoo@mcgill.ca

<sup>3</sup>Postdoctoral Fellow, Dept. of Bioresource Engineering, McGill Univ., 21,111 Lakeshore Rd., Ste. Anne de Bellevue, QC, Canada, H9X 3V9; presently, Assistant Professor, Dept. of Civil Engineering, Indian Institute of Technology, Guwahati 781039, India (corresponding author). E-mail: vipmkgoyal@gmail.com

<sup>4</sup>Agrometeorologist, Caribbean Institute of Meteorology and Hydrology, Husbands, St. James BB23006, Barbados. E-mail: atrotman@cimh.edu.bb

Note. This manuscript was submitted on October 10, 2012; approved on April 23, 2013; published online on April 25, 2013. Discussion period open until April 1, 2014; separate discussions must be submitted for individual papers. This paper is part of the *Journal of Irrigation and Drainage Engineering*, Vol. 139, No. 11, November 1, 2013. © ASCE, ISSN 0733-9437/2013/11-922-932/\$25.00.





Fig. 1. Locations of climate stations

Jamaica experienced below normal rainfall, causing significant losses in the agricultural sector. The Jamaican government had to respond to significant losses in the sugar sector by offering the sector a US\$100 million assistance package in 1997. Subsequently, between October 1999 and March 2000, rainfall was less than 25% of normal in some places, resulting in crop losses of approximately US\$6 million. Hence, drought index information is essential for better planning for drought impacts and will allow for the introduction of mitigation measures by the agricultural sector.

The Standardized Precipitation Index (SPI) is the most internationally used drought indicator and is used to assess anomalous and extreme precipitation (Andreau et al. 2007). This index was developed by McKee et al. (1993) and has been applied extensively in many parts of the world, including the United States (Hayes et al. 1999), Australia (Barros and Bowden 2008), Europe (Cancelliere et al. 2007), and Africa (Ntale and Gan 2003). The index is computationally simple and time-flexible, meaning that it can be developed over different time scales (Bonaccorso et al. 2003; Cancelliere et al. 2007; Guttman 1998; Mendicino et al. 2008). The Normalized Difference Vegetative Index (NDVI) is an index that is used to measure and monitor plant growth and vegetation cover, and is derived from remote sensing measurements (USGS 2010). The NDVI is calculated from the red and near-infrared reflectance from vegetation, which is measured by satellite.

The NDVI has also been found to correlate to monthly mean soil moisture values in previous studies (Farrar et al. 1994). However, the NDVI was found to better represent soil moisture in dry years than wet years because of high soil moisture availability. The NDVI did not correlate well with soil moisture for brush species in rangeland or trees in forest land; however, it responded well to changes in soil moisture for agricultural and pasture lands (Narasimhan et al. 2005). Narasimhan et al. (2005) also found that the NDVI provides a good representation of soil moisture and can be used as a good agricultural drought indicator.

The primary objectives of this study are: (1) to evaluate the suitability of both the SPI and the NDVI in reflecting water stressed conditions for three agricultural areas in Jamaica; (2) to determine the relationship between the index values and available soil moisture on a monthly basis; and (3) to calculate monthly irrigation demands for the three study sites for vegetables and sugarcane.

## Materials and Methods

### Descriptions of Study Sites

Three sites were used in this study (Fig. 1): Savanna-la-Mar in the parish of Westmoreland, Beckford Kraal in the parish of Clarendon, and Serge Island in the parish of St. Thomas. These sites were selected because there are historical rainfall data spanning a minimum period of 30 years. Each site has distinctly different soil characteristics and farming practices. The soils have great spatial variability within all three parishes. For the purposes of this research, the soil that dominated the 500-m radius of each climate station was used. The basic characteristics of each site are described in the following.

Savanna-la-Mar has three distinct growing seasons. These growing seasons range from September to December, January to April, and May to August. Crops such as Irish potatoes (*Solanum tuberosum*), carrots (*Daucus carota*), tomatoes (*Lycopersicon esculentum*), sweet peppers (*Capsicum annum*), cauliflower (*Brassica oleracea botrytis*), and cabbage (*Brassica oleracea capitata*) are grown during the months of September to April in two rotations of three to four months each, whereas perennial crops such as pineapples (*Ananas comosus*), papayas (*Carica papaya*), plantains (*Musa paradisiaca*), and bananas (*Musa sapientum*) are harvested during the summer months of May to August (W. Mitchell, personal communication, 2010). Loam soil is the dominant soil type.

At Beckford Kraal, vegetable crops are rotated three to four times throughout the entire year, despite seasonal variations in rainfall (P. Stone, personal communication, 2010). The crops grown are callaloo (*Amaranthus viridis*), carrots, cauliflower, lettuce, pak-choy (*Brassica rapa var. chinensis*), cabbage, and pumpkins (*Cucurbitapepo*). Clay is the dominant soil type. At Serge Island, there are also multiple rotations of the vegetable crops throughout the entire year (L. Hemans, personal communication, 2010). These crops include carrots, tomatoes, pumpkins, and cabbages. Sandy loam is the dominant soil type. Sugarcane can be grown year-round at all three locations. The typical harvesting time can range anywhere from December to April.

### Determination of Soil Moisture

A conceptual soil moisture model based on the water balance was used for this research (Chin 2006). The soil moisture model is

based on a monthly accounting of the water balance. This model is based on the assumption that as the amount of water within the soil column decreases, the rate at which it can be removed from the soil also decreases. As this model determines available soil moisture, it is inherently bound between the field capacity (FC), and the wilting point (WP) of the soil. The model splits the soil column into two layers: upper and lower soil layers. The following equations are used to account for available soil moisture:

$$L_s = \min[S'_s, (ET_p - P)] \quad (1)$$

$$L_u = (ET_p - P - L_s) \frac{S'_u}{w_{ac}}, \quad \text{provided that } L_u \leq S'_u \quad (2)$$

where  $L_s$  = moisture loss from the surface layer (mm);  $S'_s$  = available moisture in the surface layer at the start of the month (mm);  $ET_p$  = potential evapotranspiration (mm);  $P$  = monthly precipitation (mm);  $L_u$  = moisture loss from the underlying soil (mm);  $S'_u$  = available moisture stored in the underlying soil at the start of the month (mm); and  $w_{ac}$  = available water capacity of the soil (mm) (Chin 2006). Values of FC and WP were obtained from the study by Schwab (1993), and are generic values based on soil textural information. Here, all values of available soil moisture listed in this paper are soil moisture depth per meter of soil. The available soil water capacity values are 170 mm for Savanna-la-Mar, 230 mm for Beckford Kraal, and 120 mm for Serge Island.

The simulation was started on November 1 for both the Savanna-la-Mar and Serge Island sites, because this is at the peak of the wet season. Therefore, it was assumed that the soil was at FC during this time. However, for the Beckford Kraal site, the simulation was started on June 1 because during the period from September to November through the period of 1970 to 1980, a deficit was observed between total monthly precipitation and potential evapotranspiration, and it would not have been reasonable to assume that the soil was at FC during these months. As such, it was assumed that the soil was at FC on June 1.

### Determination of Irrigation Requirements

The irrigation requirement was derived from the relationship modified from Savya and Frenken (2002):

$$IR = ET_c - (Pe + Ge + SWa) + LR_{mm} \quad (3)$$

where  $IR$  = irrigation requirement (mm);  $ET_c$  = monthly crop evapotranspiration (mm);  $Pe$  = monthly effective rainfall (mm);  $Ge$  = monthly groundwater contribution from water table (mm);  $SWa$  = plant available water stored in the soil at the end of previous month (mm); and  $LR$  = monthly leaching requirement (mm).

The determination of each of the preceding parameters will be discussed in the following.

### Crop Evapotranspiration

Class A pan evaporation was used to determine crop evapotranspiration for both sugarcane and vegetables. For the purposes of this study, vegetables represent cabbages, carrots, cauliflower, and lettuce. Ten year average monthly pan evapotranspiration values were available for each area; these were used along with the relevant pan coefficient and crop coefficient ( $K_c$ ) values to determine  $ET_c$  values for each crop. The  $K_c$  values used in this study were for crops in humid climates [minimum relative humidity ( $RH_{min}$ ) > 70%]. A graphical method was used (Allen et al. 1998) to determine the  $K_c$  values for humid climates. There was a jump from a  $K_c$  value of 0.3 in April to 0.95 in May for sugarcane, because April is the end of

the harvesting period and May is the beginning of the initial growth period. A pan coefficient of 0.85 was used and chosen based on data published by Allen et al. (1998). No relative humidity data were easily available for any of the climatological stations used in this study. However, relative humidity data were available for a few other stations in other parts of the island. The relative humidity for all other stations remained above 70% for all months of the year. In addition, a light wind speed of <2 m/s was assumed for all three areas, because this is a reasonable assumption to make for areas with a high relative humidity (Allen et al. 1998). It is for these reasons that the pan coefficient of 0.85 was deemed suitable for use in all three areas for all months of the year.

### Effective Rainfall

To determine the irrigation requirements for each study area, the rainfall levels associated with 80 and 90% probability of exceedance were found for each month for each station. These values were chosen based on recommendations by Savya and Frenken (2002). The rainfall associated with the 90% probability of exceedance was used for the months of December to April, because these months experience the least amount of rainfall. However, for the months of May to November, the rainfall values with an 80% probability of exceedance were used. To determine the monthly effective rainfall, the following empirical equation was used (Bos et al. 2009):

$$Pe = f \times (1.253P^{0.824} - 2.935) \times 10^{0.001ET_c} \quad (4)$$

where  $Pe$  = effective precipitation per month (mm/month);  $f$  = correction factor that depends on the depth of the irrigation water supplied per irrigation application;  $P$  = precipitation per month (mm/month); and  $ET_c$  = total crop evapotranspiration per month (mm/month).

The value of  $f$  is calculated as follows (Bos et al. 2009):

$$f = 0.133 + 0.201 \ln D_a \quad \text{if } D_a < 75 \text{ mm/application} \quad (5)$$

and

$$f = 0.946 + 7.3 \times 10^{-4} \times D_a \quad \text{if } D_a \geq 75 \text{ mm/application} \quad (6)$$

where  $D_a$  = depth of water supplied per irrigation application. Here,  $D_a$  was assumed to be 170 mm for Savanna-la-Mar, 230 mm for Beckford Kraal, and 120 mm for Serge Island.

### Groundwater Contribution, Available Plant Water, and Leaching Requirement

Groundwater table data were collected by the Water Resources Authority of Jamaica for different locations within the island, and are available on their website at [www.wra.gov.jm](http://www.wra.gov.jm). However, no recent water table levels that were part of a historical time series were available for the study areas. As a result, it was deemed necessary to ignore groundwater contribution to crop water requirements.

For the irrigation requirements, the available water of the soil was determined based on the results from the correlation and regression analysis of the SPI and calculated soil moisture. The irrigation requirements were determined for the "moderately dry" and "severely dry" SPI categories by using the soil moisture values obtained from the regression analysis of the relevant SPI values during the months of the year for which good coefficients of regression were obtained.

To manage the high salt conditions in the root zone, extra water can be used for irrigation in a process called leaching (Savya and Frenken 2002). The leaching requirement is the excess amount of irrigation water used for this process and depends on the irrigation

water salinity and the tolerance of the crop to salinity. Because the soils in the study areas do not have high salinities, the leaching requirement was ignored for the purposes of these calculations.

### Development of NDVI

The NDVI has been shown to be a good indicator of vegetation health because chlorophyll absorbs broadband red wavelengths and reflects near-infrared wavelengths (Rogers et al. 2009). The NDVI is calculated from the red and near-infrared reflectance from the vegetation, measured by satellite, and is calculated by the following ratio:  $(\text{near-infrared} - \text{red}) / (\text{near-infrared} + \text{red})$  (Samson 1993).

The NDVI for Jamaica was obtained from the United States National Oceanic Atmospheric Administration (NOAA) advanced very high radiometry resolution (AVHRR) Landsat imagery at a 250-m spatial resolution over 16-day composites. The NDVI values were obtained directly from the vector data sets produced by NOAA and extracted over a 500-m radius from the rain gauge station for the period from 2000 to 2008. A 500-m radius was selected because this was deemed to be a conservative approximation of the minimum area that would be affected by a rainfall event. The pixel values for the NDVI were averaged over this 500-m radius for each 16-day composite and tabulated. Each of these 16-day composites underwent a time-weighted average smoothing procedure to obtain the NDVI for each month.

### Seasonality Analysis of the Soil Moisture and NDVI Time Series

The distinct possibility exists that the seasonal component of a time series may lead to concerns such as covariance and autocorrelation within the time series; therefore, model may yield inaccurate results with regression analysis (Ji and Peters 2003; Thompstone et al. 1985; Wang et al. 2007; Weissling and Xie 2009). There are several ways to address this seasonality issue. Ji and Peters (2005) used dummy variables to account for seasonality effects within their regression analysis. However, the issue of seasonality can also be addressed by removing the seasonal components from the time series. Therefore, in this study, the correlation and regression analysis were performed with the deseasonalized soil moisture and deseasonalized NDVI time series. The deseasonalization process was conducted on the 16-day composites of the NDVI, before they were smoothed into one-month composites, and on the monthly soil moisture.

The seasonal time series for the NDVI was first determined by calculating the average 16-day composite NDVI for each 16-day period over the nine-year time period. A three-point moving average was calculated for each 16-day time period to obtain a seasonal value for each 16-day time period. To determine the deseasonalized time series, this seasonal value was subtracted from the value of each time period over the entire nine-year time series. The same procedure was used for the soil moisture, with the difference that each "season" was one month. The process is described as follows, which was modified from Weissling and Xie (2009):

$$TS_{ds} = TS - TS_{sm} \quad (7)$$

$$TS_{sm} = \left( \sum_{j=t-1}^{t+1} \frac{\sum_{i=1}^n TS}{n} \right) / 3 \quad (8)$$

where  $TS_{ds}$  = 16-day deseasonalized mean in the NDVI time series (or the monthly deseasonalized mean in the soil moisture time series);  $TS$  = 16-day mean in a time series;  $TS_{sm}$  = three-point

smoothed 16-day mean in a time series;  $j$  = value immediately before and after an event value at time  $t$ ; and  $n$  = number of years for which the deseasonalized time series is computed. Ideally, long-term means are considered appropriate for deseasonalizing time series (Wang et al. 2007; Weissling and Xie 2009). However, this method has been successfully applied to much shorter NDVI time series than in this study (Weissling and Xie 2009); for that reason, the nine-year NDVI time series in this study is considered to be appropriate.

### Standardized Precipitation Index

The SPI is a meteorological index based solely on precipitation (McKee et al. 1993). The index is developed by using monthly precipitation data, which is ideally continuous over at least 30 years. The SPI can be developed over different time scales, such as one, three, six, 12, 24, and 48 months. A three-month SPI for a particular month takes into account the precipitation for the month in question and that of the two previous months. The precipitation data sets are applied to a Gamma distribution function (McKee et al. 1993). This allows for the establishment of a relationship between probability and precipitation, leading to the calculation of a normally distributed probability density with a mean of zero and SD of unity (McKee et al. 1993). Thus, negative values of the SPI represent dry conditions, whereas positive values represent wet conditions.

For the purposes of this research, the SPI was obtained by using a programming tool developed by the United States National Drought Mitigation Center (2006). The SPI was developed for the one, three, six, nine, and 12 month periods for the sites. The monthly rainfall data were obtained directly from the Meteorological Service of Jamaica from 1971–2008 for all three sites. To correlate the SPI to soil moisture, the one-month and three-month SPI for each month were correlated to the concurrent monthly soil moisture. This was done over the entire 38-year time series. For example, the three-month SPI for March 1971 was compared to the monthly soil moisture for March 1971. Like the NDVI, attempts were also made to lag the soil moisture by one and two months to determine whether the correlation results would improve.

### Correlation and Regression Analysis

A bivariate correlation and regression analysis was conducted between monthly NDVI and soil moisture values. The same analysis was also conducted on the soil moisture and the one-month and three-month SPI values. A least-squares regression analysis was conducted on the two analyses, with correlation coefficients reported within a 5% significance level. It was found through regression analyses that not all relationships were first order (linear). Cross-correlation (lag) analysis was conducted for the NDVI and soil moisture time series by lagging the NDVI for one and two months. The NDVI was compared to soil moisture on a monthly basis, not as a continuous time series. This was done because the relationship between vegetation and soil moisture can change from month to month (Wang et al. 2007). The SPI and available soil moisture were also compared on a month-by-month basis, because the relationship between precipitation (and thus SPI) and soil moisture is different for each month. In general, in wet months, small precipitation events can lead to the soil reaching FC, whereas in dry months, much larger precipitation events, or more frequent smaller events, would be needed for the soil to reach FC.

The values of soil moisture were obtained by using the regression coefficients obtained from the regression analyses. To determine the predicted value of the soil moisture from the regression



model, the seasonal value simply needs to be added to this predicted deseasonalized value. These relationships were used to determine soil moisture values for the different categories of the SPI, as used by the NOAA National Climatic Data Center (NCDC). These categories range from “severely dry” to “exceptionally moist.” This paper includes soil moisture values relating to the categories “severely dry” to “near normal.” Tables 1–3 include soil moisture values from the “near normal” to “exceptionally moist” categories. All values are reported within a 5% level of significance.

**Table 1.** Values of Available Soil Moisture for Various Classifications of Positive SPI Values for Savanna-la-Mar

Water availability	SPI-3					
	values	February	March	April	May	June
Near normal	−0.5	3.3	1.1	1.5	5.8	6.6
	0.5	7.0	4.7	5.6	12.1	12.2
Abnormally moist	0.51	7.1	4.7	5.7	12.2	12.3
	0.79	8.1	6.2	7.1	13.9	13.9
Moderately moist	0.8	8.1	6.2	7.2	14.0	13.9
	1.29	10.0	9.1	10.0	17.0	16.7
Very moist	1.3	10.0	9.2	10.1	FC	16.8
	1.59	11.1	11.2	12.0	FC	FC
Extremely moist	1.6	11.1	11.2	12.0	FC	FC
	1.99	12.6	14.2	14.8	FC	FC
Exceptionally moist	2	12.6	14.3	14.9	FC	FC

Note: All values are in mm.

**Table 2.** Values of Available Soil Moisture for Various Classifications of Positive SPI Values for Beckford Kraal

Water availability	SPI-3						
	value	February	March	April	May	June	September
Near normal	−0.5	7.7	6.1	6.0	9.7	8.4	9.0
	0.5	12.9	11.8	11.7	17.3	14.9	16.5
Abnormally moist	0.51	12.9	11.9	11.8	17.3	15.0	16.6
	0.79	14.4	13.4	13.4	19.5	16.8	18.7
Moderately moist	0.8	14.5	13.5	13.4	19.5	16.9	18.8
	1.29	17.0	16.3	16.2	FC	20.1	22.5
Very moist	1.3	17.1	16.3	16.3	FC	20.1	22.5
	1.59	18.6	18.0	17.9	FC	22.0	FC
Extremely moist	1.6	18.6	18.0	18.0	FC	22.1	FC
	1.99	20.7	20.2	20.2	FC	FC	FC
Exceptionally moist	2	20.7	20.3	20.2	FC	FC	FC

Note: All values are in mm.

**Table 3.** Values of Available Soil Moisture for Various Classifications of Positive SPI Values for Serge Island

Water availability	SPI category	SPI-1			
		value	February	March	April
Near normal	−0.5 to +0.5	0.5	3.8	2.5	8.8
		−0.5	1.2	0.3	4.5
Abnormally moist	0.51 to 0.79	−0.51	1.1	0.3	4.4
		0.79	4.9	3.4	10.0
Moderately moist	0.80 to 1.29	0.8	4.9	3.4	10.0
		1.29	6.9	5.1	12.0
Very moist	1.30 to 1.59	1.3	7.0	5.1	12.0
		1.59	8.3	6.3	FC
Extremely moist	1.60 to 1.99	1.6	8.3	6.4	FC
		1.99	10.3	8.1	FC
Exceptionally moist	2.00 and above	2	10.4	8.1	FC

Note: All values are in mm.

## Correlation and Regression Analysis between SPI and Soil Moisture for El Niño Years

As mentioned previously, the SPI best represents soil water values during the dry months of the year. Also mentioned previously, El Niño years have been responsible for many of the droughts Jamaica has experienced in the past, resulting in drier than normal conditions during the dry seasons. With this in mind, the El Niño years between 1970 and 2008 were selected and the SPI values were correlated to the values of the soil water on a monthly basis only for these El Niño years. This was done using three variations. The first variation used all of the El Niño years (1972–1973, 1976–1977, 1982–1983, 1986–1987, 1991–1992, 1994–1995, 1997–1998, 2002–2003, and 2006–2007). The second variation used the first year in each of the two-year El Niño occurrences. The third variation used the second year in each of the two-year El Niño occurrences. For example, the third variation used the years 1973, 1977, and 1983. This was done because the effects of El Niño would most likely be felt in the dry months of the second year of the El Niño occurrence, but not in the dry months of the previous years, because the effects of El Niño start to manifest themselves nearer to the end of the first year. As such, isolating the years allows the results to reflect this.

## Results

### Correlation and Regression Analysis for NDVI and Soil Moisture

By comparing NDVI and soil moisture on a monthly basis, it was found by using the correlation analysis that the coefficient of determination ( $r^2$ ) was  $\geq 0.7$  in the months of January and March for Savanna-la-Mar, January and April for Beckford Kraal, and January and February for Serge Island. The  $r^2$  regression coefficients for each of these months are shown in Table 4.

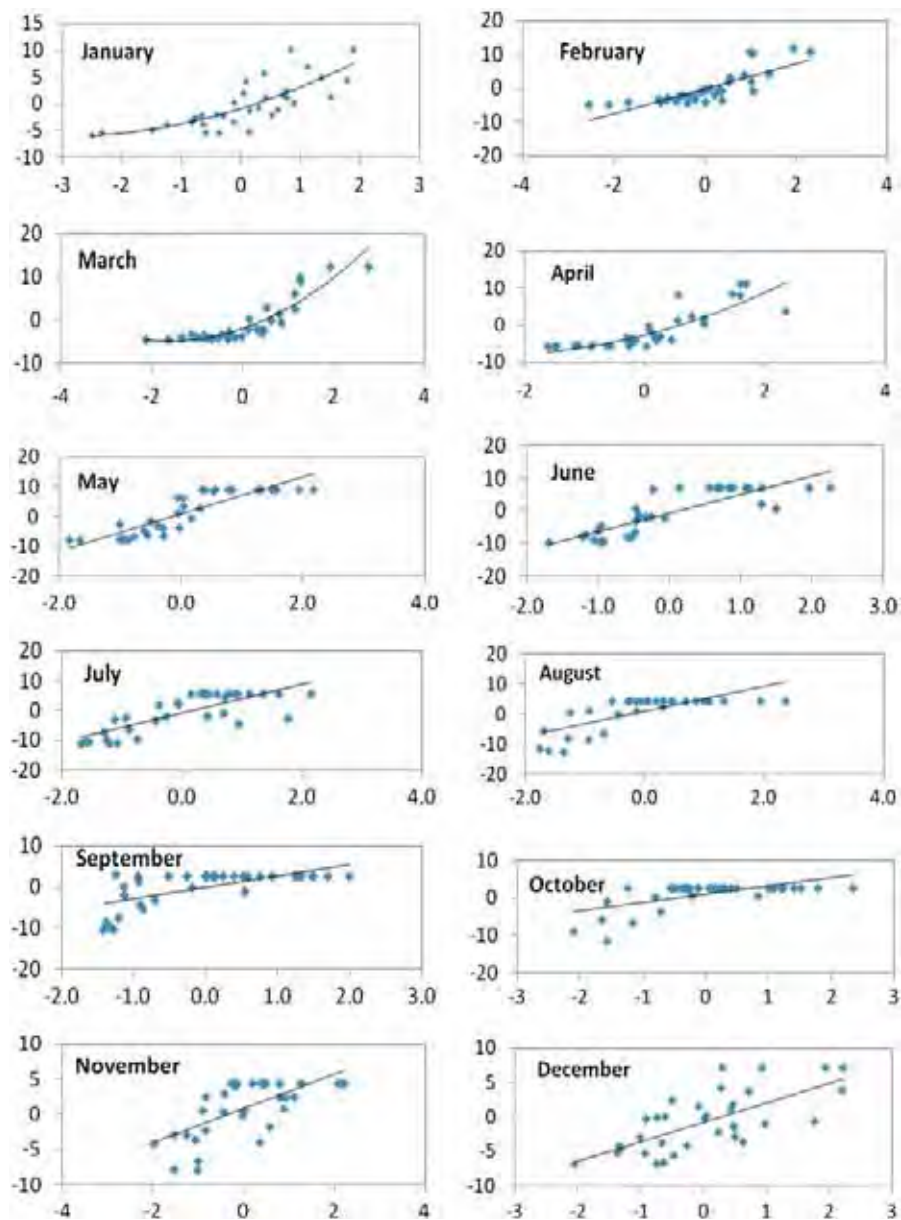
### Correlation and Regression Analysis for SPI and Available Soil Moisture

The three-month SPI (SPI-3) showed reasonable  $r^2$  regressions ( $\geq 0.7$ ) for the months of February to June for Savanna-la-Mar, and for the months of February to June and September for Beckford Kraal. Scatter plots for deseasonalized soil moisture values versus SPI values for Savanna-la-Mar are shown in Fig. 2. However, the one-month SPI (SPI-1) had the best regression fits for Serge Island, but reasonable regression fits were only observed for the months of February, March, and August. Table 5 shows the  $r^2$  coefficients for each month. The coefficients for the SPI-3 values and available soil moisture are reported for Savanna-la-Mar and Beckford Kraal unless otherwise indicated. The coefficients for SPI-1 and available soil moisture are shown for Serge Island.

Although the months of May and June have  $r^2$  coefficients of 0.7, they were not listed as months for which reasonable regression

**Table 4.**  $r^2$  Regression Coefficients for NDVI and Available Soil Moisture for Relevant Months

Location	Month	$r^2$
Savanna-la-Mar	January	0.73
	March	0.73
Beckford Kraal	January	0.75
	April	0.71
Serge Island	January	0.93
	February	0.74



**Fig. 2.** Scatter plots for deseasonalized soil moisture values versus SPI values for Savanna-la-Mar (all SPI values are SPI-3 values except for January, which is SPI-1)

fits were obtained. The reason for this is that autocorrelation was found between the residuals for these months. Also, for both Savanna-la-Mar and Beckford Kraal, the  $r^2$  coefficients were higher for SPI-1 for the month of January than SPI-3, so these are listed in the table.

### **SPI and Available Soil Moisture**

As shown in Table 5, the SPI only had good correlations to available soil moisture in particular months. One possible reason is that

in dry months, the changes in precipitation are better reflected in soil moisture. May, although considered to be a wet month, had good correlations for both Savanna-la-Mar and Beckford Kraal. This may be because it immediately follows the driest months of the year so the soil would likely not be at FC and would still respond to increases in precipitation. Therefore, during May, the increase in rainfall is accompanied by increasing soil moisture, resulting in good correlations.

Tables 6–8 show available soil moisture values for plants based on the categories of the SPI used by the NOAA NCDC for each

**Table 5.**  $r^2$  Regression Coefficients for SPI and Soil Moisture

Location	January	February	March	April	May	June	July	August	September	October	November	December
Savanna-la-Mar	0.6 <sup>a</sup>	0.7	0.8	0.7	0.8	0.7	0.6	0.5	0.4	0.4	0.4	0.5
Beckford Kraal	0.5 <sup>a</sup>	0.7	0.7	0.7	0.8	0.7	0.6	0.5	0.7	0.5	0.3	0.3
Serge Island	0.6	0.7	0.7	0.4	0.7 <sup>b</sup>	0.7 <sup>b</sup>	0.6	0.7	0.5	0.6	0.6	0.6

<sup>a</sup> $r^2$  regression coefficients for SPI-1 relationships.

<sup>b</sup>Autocorrelation found between the residuals in these months.



**Table 6.** Available Soil Moisture Values of Plants for the SPI Categories for Savanna-la-Mar

Water availability	SPI-3					
	values	February	March	April	May	June
Near normal	0.5	70	47	56	121	122
	-0.5	33	11	15	58	66
Abnormally dry	-0.51	32	11	15	58	65
	-0.79	22	5	6	40	50
Moderately dry	-0.8	22	4	6	40	49
	-1.29	3	0	0	9	21
Severely dry	-1.3	3	0	0	0.8	21
	-1.59	0	0	0	0	4
Extremely dry	-1.6	0	0	0	0	4
	-1.99	0	0	0	0	0
Exceptionally dry	-2	0	0	0	0	0

Note: All values are in mm.

**Table 7.** Available Soil Moisture Values of Plants for the SPI Categories for Beckford Kraal

Water availability	SPI-3						
	value	February	March	April	May	June	September
Near normal	0.5	129	118	117	173	149	165
	-0.5	77	61	60	97	84	90
Abnormally dry	-0.51	76	61	60	96	84	89
	-0.79	62	45	44	75	66	68
Moderately dry	-0.8	61	44	43	74	65	67
	-1.29	36	16	16	37	33	31
Severely dry	-1.3	35	16	15	36	33	30
	-1.59	20	0	0	14	14	8
Extremely dry	-1.6	20	0	0	14	13	07
	-1.99	0	0	0	0	0	0
Exceptionally dry	-2	0	0	0	0	0	0

Note: All values are in mm.

**Table 8.** Available Soil Moisture Values of Plants for the SPI Categories for Serge Island

Water availability	SPI-1 value	February	March	August
Near normal	0.5	38	25	88
	-0.5	12	3	45
Abnormally dry	-0.51	11	3	44
	-0.79	6	0	32
Moderately dry	-0.8	6	0	31
	-1.29	0	0	8
Severely dry	-1.3	0	0	8
	-1.59	0	0	0
Extremely dry	-1.6	0	0	0
	-1.99	0	0	0
Exceptionally dry	-2	0	0	0

Note: All values are in mm.

location and for each relevant month, for negative values of the SPI. The categories that represent water scarcity have been defined by NOAA NCDC as “near normal,” “abnormally dry,” “moderately dry,” “severely dry,” “extremely dry,” and “exceptionally dry.” Each of these categories is defined by a range in SPI values. For example, the “near normal” category is defined by a range in SPI values of -0.5 to 0.5, whereas the “abnormally dry” category is defined by a range of -0.79 to -0.51. Because these are the predefined categories and definitions used in the United States, the same categories will be applied within this research.

The values are bounded at the lower limit at 0 mm and at the upper limit at the available water capacity of the soil. The curves

**Table 9.**  $r^2$  Coefficients for the SPI and Soil Moisture in Selected El Niño Years

Climate station	Month	$r^2$ correlation
Savanna-la-Mar	May	0.79 <sup>a</sup>
	June	0.85
	July	0.85
	September	0.73
Beckford Kraal	December	0.72
	January	0.87 <sup>a</sup>
	May	0.91
Serge Island	June	0.76
	March	0.67
	May	0.76 <sup>a</sup>
	June	0.79 <sup>a</sup>
	July	0.81 <sup>a</sup>
	October	0.69 <sup>a</sup>
	December	0.75 <sup>a</sup>

<sup>a</sup> $r^2$  regression coefficients for SPI-1 relationships.

cannot represent these boundary conditions, so the values had to be forcibly bounded at the lower and upper limits of the soil moisture. For each category of water scarcity (such as “near normal”), the available water of plants is shown for the lower and upper limit of that category. Therefore, in Table 6, 0.5 represents the upper limit of the “near normal” category, and the available water of plants is shown for this SPI value, as is the lower limit of that category.

For both Savanna-la-Mar and Beckford Kraal, an available water content of zero for soil occurs the earliest during March and April, meaning that it occurs at the lower boundary of moderately dry/upper boundary of the “severely dry” category, with corresponding SPI values of -1.29 and -1.30, respectively. This supports the fact that these are the driest months (Fig. 3 shows monthly average rainfall for each three locations). Interestingly, for Savanna-la-Mar, it occurs the latest, in the month of June (the lower boundary of the “extremely dry” category with a corresponding SPI value of -1.99), which suggests that soil moisture during this month exceeds that of May. For Beckford Kraal, an available water content for soil of zero occurs at the same point (the lower boundary of the “extremely dry” category) in May. However, there is very little difference between the soil moisture values at the upper end of this category (SPI value of -1.60), with a 10-mm difference between May and June. For Serge Island, the wettest month represented in Table 8 is August, with March as the driest month again. An available water content for soil of zero occurs at the upper end of the “moderately dry” category in March and at the lower bound of the “severely dry” category in August. An available water content for soil of zero occurs at less severe water scarcity conditions in March for Serge Island than for either Savanna-la-Mar or Beckford Kraal. It is likely that the soil type is the primary reason for this. The influence of soil type on the relationship between SPI values and available soil moisture is discussed later in this paper.

### Effects of ENSO on the Relationship between SPI and Soil Water

The results of the third variation (i.e., that includes the second year in each of the two-year El Niño occurrences) are better than those of the first or second, supporting the fact that the dry months of the second years are affected more than those in the first years. Table 9 shows all of the months for which  $r^2$  values of 0.7 or greater were achieved.

Serge Island has the most months with  $r^2$  correlations  $>0.7$ , far surpassing the number of months with good correlations when all the years from 1970 to 2008 were considered. This suggests that the dryer the year, the stronger the relationship between the SPI and soil moisture conditions. Also, SPI-1 had the best correlations, with only March having a better correlation with SPI-3. The Beckford Kraal results did not improve with just the El Niño years, because there were fewer months (only three instead of five), with  $r^2$  values  $>0.7$ . However, for Savanna-la-Mar, both September and December showed good correlations, which is not the case when all of the years are taken into consideration.

These tests suggest that the dryer the conditions, the stronger the relationship between the SPI values and soil water values. This was supported by the results for Serge Island and Savanna-la-Mar, although not by the results for Beckford Kraal. However, it is probable that this is because Beckford Kraal has a clay soil, which would not represent the changes in soil water as readily.

### Applicability of Results to Determining Irrigation Requirements

The SPI values can be used to determine monthly soil moisture values during water scarcity conditions. The information provided in Tables 6–8 can be used based on the drought condition that is being experienced. These soil moisture values can be used to determine irrigation requirements during water scarcity conditions. Irrigation is predominantly required during the dry months of the year. There are limitations because there are no good correlations for the wet months of the year. If there is water scarcity during these months, this tool would not be applicable. However, it provides a means of determining the irrigation requirements during the months of the year in which the highest irrigation dependency exists.

Irrigation requirements for the lower bounds of the “moderately dry” and “severely dry” SPI categories have been calculated for the three sites and are shown in Tables 10–12. The information in these tables is meant to be used for planning purposes. For both Savanna-la-Mar and Serge Island, available soil water values of 0 mm were experienced during the “moderately dry” periods for March and April, and February and March, respectively. As a result, the irrigation requirements for these months are the same, regardless of the drought intensity. An innovative tool now exists for determining irrigation requirements during water scarcity conditions.

## Discussion

### NDVI and Available Soil Moisture

The NDVI only had reasonable correlations for months during the driest period of the year (January and March for Savanna-la-Mar, January and April for Beckford Kraal, and January and February

**Table 10.** Irrigation Requirements for the Moderately and Severely Dry Categories of Drought for Savanna-la-Mar

SPI drought category	Crop	Irrigation requirement at end of month (mm)				
		February	March	April	May	June
Moderately dry	Vegetables	121	140	122	11	52
	Sugarcane	68	37	19	39	41
Severely dry	Vegetables	121	140	122	11	52
	Sugarcane	85	48	33	81	74

**Table 11.** Irrigation Requirements for the Moderately and Severely Dry Categories of Drought for Beckford Kraal

SPI drought category	Crop	Irrigation requirement at end of month (mm)					
		February	March	April	May	June	September
Moderately dry	Vegetables	70	91	97	10	93	0
	Sugarcane	0	0	0	9	10	9
Severely dry	Vegetables	86	107	113	26	109	0
	Sugarcane	0	10	15	32	29	32

**Table 12.** Irrigation Requirements for the Moderately and Severely Dry Categories of Drought for Serge Island

SPI drought category	Crop	Irrigation requirement at end of month (mm)		
		February	March	August
Moderately dry	Vegetables	119	148	100
	Sugarcane	94	121	108
Severely dry	Vegetables	119	148	100
	Sugarcane	94	121	116

for Serge Island). Because of the limited months of good regressions between the NDVI and soil moisture, it is not recommended for use at these sites.

Although none of these sites could be classified as arid or semiarid, the fact that the dry months are the only times during which relationships exist between the NDVI and available soil moisture suggests that the vegetation response to soil moisture is far stronger during dry than wet periods in these high rainfall zones. Narasimhan et al. (2005) stated that the study site within the high rainfall zone of their study area had low correlations between NDVI and soil moisture, because the NDVI did not fluctuate much with changes to soil moisture as a result of the high annual rainfall. There are three primary factors that could have influenced the results: soil type, aridity, and vegetation.

### Effect of Soil Type

Soil types such as clay tend to retain water for much longer periods than sandy soils (because of poor drainage and high available water capacity in clay soils), changes in precipitation are not reflected in soil moisture, and therefore, the vegetation, as readily as it would in a sandy soil. In this study, the highest  $r^2$  regression coefficient was 0.93 for January in Serge Island. This island has a sandy loam, which has the smallest available water capacity of all of the soils. This is supported by Farrar et al. (1994), who reported that in their study, the soil moisture was higher in the cambisols and vertisols (soils with the highest clay content), with resulting lower NDVI correlation, than in the arenosols (soils with the highest sand content).

### Effect of Aridity

Because of the high annual rainfall that each of the sites receive, the reasonable correlations were only observed for the driest months of the year for the study locations, which supports the idea that the NDVI is really only a robust estimator of soil moisture in arid or semiarid areas.

Farrar et al. (1994) investigated the relationship between NDVI and soil moisture in semiarid Botswana, and this study also determined that the NDVI was correlated to the surface soil moisture of the concurrent month of the growing season. Wang et al. (2007) stated that in semiarid environments, NDVI changes closely with

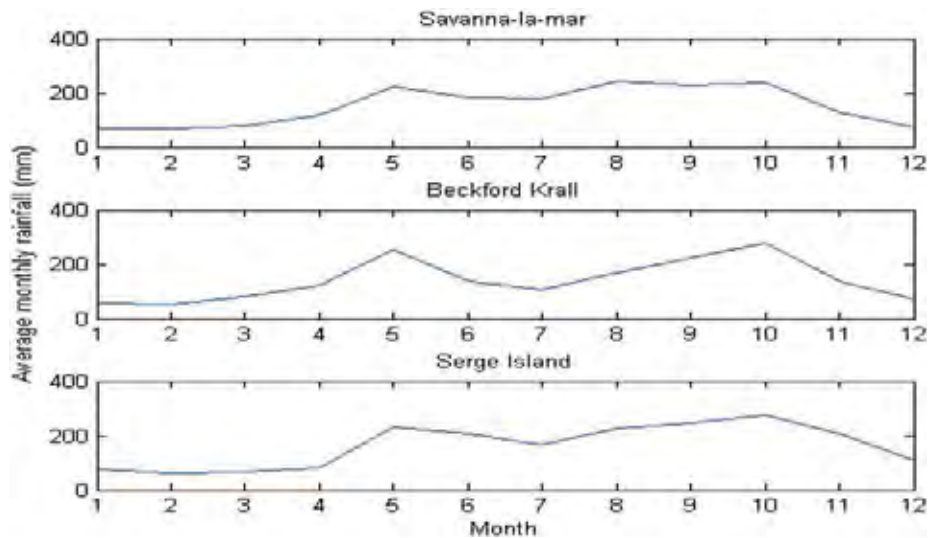


Fig. 3. Monthly rainfall hydrographs for Savanna-la-Mar, Beckford Kraal, and Serge Island

soil moisture because soil moisture is the major controlling factor for vegetation; they also found that the NDVI at humid sites takes longer (10 days) to respond to soil moisture than at arid sites (five days). In this study, the NDVI and soil moisture could not be compared statistically for the drought years 2000–2001, because two years is not sufficiently long to perform an analysis. However, a visual, and therefore qualitative, observation of the deseasonalized time series plot of the NDVI shows distinct troughs in the 2001 to 2002 period for Serge Island, the 2000 to 2002 period for Beckford Kraal, and 2001 for Savanna-la-Mar (Figs. 4–6). This qualitative examination of the data supports the idea that the dryer the conditions, the more representative the NDVI is of the moisture conditions.

### Effect of Vegetation Type

Regarding vegetation type, Narasimhan et al. (2005) mentioned that NDVI did not correlate well to soil moisture for brush species in rangeland or trees in forest land, very possibly because of deeper rooting systems. However, much stronger correlations were observed with agricultural lands and pasturelands, because they have root systems that can only extract water from shallower depths, so this type of vegetation responds quickly to changes in soil moisture. In Jamaica, typical farms are small scale (between 1 and 2 ha) [Statistical Institute of Jamaica (STATIN) 2007]. In addition, these farms are usually interspersed with natural vegetation. As a result, it was not possible at any of the three study areas to differentiate between locations that were solely agricultural and locations that were

only brushland or woodland. This also explains the generally poor correlations between NDVI and available soil moisture for the sites.

### SPI and Soil Moisture

As mentioned previously, the dominant soil type at Serge Island was a sandy loam, which has the most limited water holding capacity of all the soils in this study. As a result, the soil moisture in any particular month would have a much smaller dependence on soil moisture in the previous months (compared to a clay soil, for instance), because of this quick response. Likewise, SPI-1 does not take into account the rainfall in previous months. However, the loam and clay soils show a slower response to rainfall because they have much larger water capacities. The soil moisture conditions in a particular month would be far more dependent on soil moisture conditions in a previous month. Likewise, SPI-3 for a particular month takes into account the two previous months of rainfall. Therefore, depending on the type of soil, either SPI-1 or SP-3 may be more useful in monitoring agricultural drought. In light of this, it is understandable why SPI-1 had the best correlations for Serge Island, whereas SPI-3 had the best correlations for Savanna-la-Mar and Beckford Kraal.

Sims et al. (2002) performed a study in North Carolina to investigate the potential of the SPI for representing short-term precipitation and soil moisture variation. The authors suggested that changes in soil types could play a significant role in the relationship between SPI and soil moisture. They suggested that SPI time series that have been averaged over longer time periods would have better correlations with soil moisture in deeper soil layers.



Fig. 4. Deseasonalized NDVI time series for Savanna-la-Mar



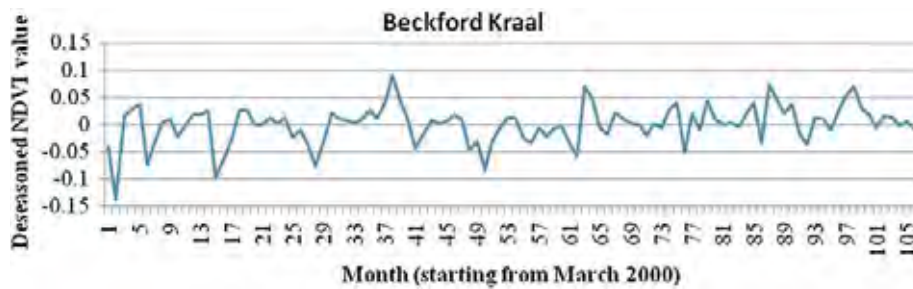


Fig. 5. Deseasonalized NDVI time series for Beckford Kraal

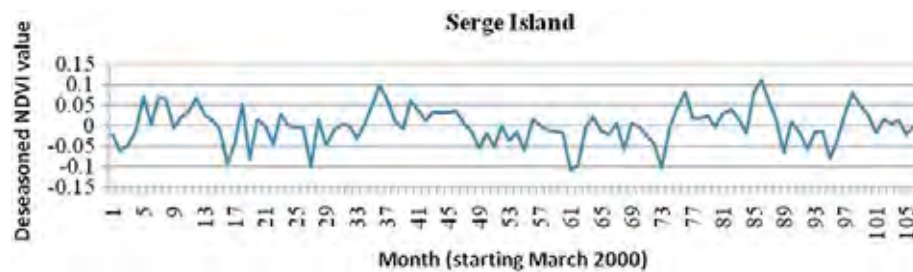


Fig. 6. Deseasonalized NDVI time series for Serge Island

Ji and Peters (2003) also found that SPI-3 is best for representing the effects of drought severity on vegetation cover. The authors suggest that this is because the impact of water deficits on vegetation is cumulative, meaning that vegetation does not respond instantaneously to precipitation. As a result, there is a time lag in the response of vegetation to precipitation. This time lag is captured by the smoothing action of SPI-3, which captures precipitation behavior over the particular month in question in addition to the two previous months. The study conducted by Sims et al. (2002) also showed that the short-term SPIs (one to three months) yielded the highest correlation between SPI and soil moisture. Serge Island had the best correlations for SPI-1. As mentioned earlier, this is most likely attributable to soil type.

## Conclusions

The applicability of both the NDVI and SPI for representing soil moisture conditions at three sites in Jamaica was evaluated. The results from this study support this conclusion, in that the only months for which the NDVI provided a suitable representation of soil moisture were the driest months of the year for all three locations. Because of the limited months for which good correlations were observed between the NDVI and soil moisture, it is not recommended for use at these sites. Soil and vegetation types play significant roles in the ability of the NDVI to represent soil moisture, and future studies involving other types of vegetation and soil might result in much better correlations.

Either SPI-1 or SP-3 was found to have reasonable  $r^2$  correlations for particular months of the year in all three study areas. SPI-3 is preferred for use at the Savanna-la-Mar and Beckford Kraal sites when planning for agricultural drought. However, SPI-1 is preferred for the Serge Island site. For the Savanna-la-Mar and Beckford Kraal locations, the months of March to June had the best correlations, whereas for the Serge Island site, the months of February, March, and August had the best correlations. A limitation to the applicability of these results to planning is the fact that the

SPI was only correlated to the soil moisture determined for particular soils. The El Niño years were isolated for the correlation and regression analysis between the SPI and soil moisture, and some good correlations were found for some of the wet months in these years, suggesting that the dryer the conditions, the stronger the relationship will be between the SPI and soil moisture.

The issue of agricultural development is a complex one in the Jamaican context, and the provision of irrigation data is one small step in increasing the viability of Jamaican farmers. Among the myriad of socioeconomic and political factors, it is becoming increasingly important that a comprehensive approach be taken to improve agricultural production within the island. The coupling of drought indices and irrigation demands is an attempt to do just that, and these results show the potential for this approach in the Caribbean.

## Acknowledgments

The authors would like to acknowledge the help of Andreas Haiduk from the Jamaica Water Resources Authority for his assistance in providing both insight and streamflow data. The Meteorological Service of Jamaica has also played a fundamental part in this research, through their provision of climatic data. The Rural Agricultural Development Agencies of Westmoreland, Clarendon and St. Thomas, as well as the Rural Physical Planning Unit, must be highly commended for their willingness to share data with the authors.

## References

- Allen, R. G., Pereira, L. S., Raes, D., and Smith, M. (1998). "Crop evapotranspiration: Guidelines for computing crop water requirements." *FAO Irrigation and Drainage Paper 56*, Food and Agriculture Organization of the United Nations, Rome.
- Andreau, J., Perez, M. A., Ferrer, J., Villalobos, A., and Paredes, J. (2007). "Drought management decision support system by means of

- risk analysis models." *Methods and tools for drought analysis and management*, G. Rossi, et al., eds., Vol. 62, Springer, Dordrecht, Netherlands.
- Barros, A. P., and Bowden, G. J. (2008). "Towards long-lead operational forecasts of drought: An experimental study in the Murray-Darling River Basin." *J. Hydrol.*, 357(3–4), 349–367.
- Bonaccorso, B., Bordi, I., Cancelliere, A., Rossi, G., and Sutera, A. (2003). "Spatial variability of drought: An analysis of the SPI in Sicily." *Water Resour. Manage.*, 17(4), 273–296.
- Bos, M. G., SpringerLink and MyiLibrary. (2009). *Water requirements for irrigation and the environment*, Springer, Dordrecht, Netherlands.
- Cancelliere, A., Di Mauro, G., Bonaccorso, B., and Rossi, G. (2007). "Stochastic forecasting of drought indices." *Methods and tools for drought analysis and management*, G. Rossi, et al., eds., Springer, Dordrecht, Netherlands.
- Chen, A. A., Falloon, T., and Taylor, M. (2005). "Monitoring agricultural drought in the West Indies." *Monitoring and predicting agricultural drought: A global study*, V. K. Boken, et al., eds., Oxford University Press, New York, 144–155.
- Chin, D. A. (2006). *Water-resources engineering*, 2nd Ed., Pearson Prentice Hall, Upper Saddle River, NJ.
- Farrar, T. J., Nicholson, S. E., and Lare, A. R. (1994). "The influence of soil type on the relationships between NDVI, rainfall, and soil moisture in semiarid Botswana (II). NDVI response to soil moisture." *Remote Sens. Environ.*, 50(2), 121–133.
- Food, and Agriculture Organization of the United Nations (FAO). (2003). *WTO agreement on agriculture—The implementation experience: Jamaica*, FAO, Rome.
- Guttman, N. B. (1998). "Comparing the Palmer drought index and the Standardized Precipitation Index." *J. Am. Water Resour. Assoc.*, 34(1), 113–121.
- Hayes, M. J., Wilhite, D. A., Svoboda, M., and Vanyarkho, O. (1999). "Monitoring the 1996 drought using the Standardized Precipitation Index." *Bull. Am. Meteorol. Soc.*, 80(3), 429–438.
- Ji, L., and Peters, A. J. (2003). "Assessing vegetation response to drought in the northern Great Plains using vegetation and drought indices." *Remote Sens. Environ.*, 87(1), 85–98.
- Ji, L., and Peters, A. J. (2005). "Lag and seasonality considerations in evaluating AVHRR NDVI response to precipitation." *Photogramm. Eng. Remote Sens.*, 71(9), 1053–1061.
- McKee, T. B., Doesken, N. J., and Kleist, J. (1993). "The relationship of drought frequency and duration to time scales." *Preprints, 8th Conf. on Applied Climatology*, Cornell University, Ithaca, NY, 179–184.
- Mendicino, G., Senatore, A., and Versace, P. (2008). "A groundwater resource index (GRI) for drought monitoring and forecasting in a Mediterranean climate." *J. Hydrol.*, 357(3–4), 282–302.
- Meteorological Service of Jamaica. (2009). *Climate: Temperature*, (<http://www.metservice.gov.jm/temperature.asp>) (Sep. 10, 2010).
- Narasimhan, B., Srinivasan, R., Arnold, J. G., and Di Luzio, M. (2005). "Estimation of long-term soil moisture using a distributed parameter hydrologic model and verification using remotely sensed data." *Trans. ASAE*, 48(3), 1101–1113.
- Ntale, H. K., and Gan, T. Y. (2003). "Drought indices and their application to East Africa." *Int. J. Climatol.*, 23, 1335–1357.
- Planning Institute of Jamaica (PIOJ). (2008). *Economic and social survey Jamaica: 2008*, PIOJ, Kingston, Jamaica.
- Richards, J., Madramootoo, C. A., and Goyal, M. K. (2013). "Determining irrigation requirements for vegetables and sugarcane in Jamaica." *Irrig. Sci.*, (in press).
- Rogers, J. C., Murrain, A. W., and Cooke, W. H. (2009). "The impact of Hurricane Katrina on the coastal vegetation of the Weeks Bay Reserve, Alabama, from NDVI data." *Estuaries Coasts*, 32(3), 496–507.
- Samson, S. A. (1993). "Two indices to characterize temporal patterns in the spectral response of vegetation." *Photogramm. Eng. Remote Sens.*, 59(4), 511–517.
- Savya, A., and Frenken, K. (2002). "Crop water requirements and irrigation scheduling." *Irrigation Manual*, K. Mudima, M. Chitima and L. Tirivamwe, eds., Vol. 4, FAO Sub-Regional Office, East and Southern Africa.
- Schwab, G. O. (1993). *Soil and water conservation engineering*, 4th Ed., Wiley, New York.
- Sims, A. P., Niyogi, D. D. S., and Raman, S. (2002). "Adopting drought indices for estimating soil moisture: A North Carolina case study." *Geophys. Res. Lett.*, 29, 24-1–24-4.
- Statistical Institute of Jamaica (STATIN). (2007). *Census of agriculture 2007—Preliminary rep.*, STATIN, Kingston, Jamaica.
- Thompstone, R. M., Hipel, K. W., and McLeod, A. I. (1985). "Forecasting quarter-monthly river flow." *J. Am. Water Resour. Assoc.*, 21(5), 731–741.
- United States National Drought, and Mitigation Center. (2006). "Monitoring drought—The standardized precipitation index." (<http://drought.unl.edu/monitor/spi/program/spiprogram.htm>) (Nov. 25, 2010).
- USGS. (2010). "What is NDVI?" (<http://ivm.cr.usgs.gov/whatndvi.php>) (May 23, 2010).
- Wang, X., Xie, H., Guan, H., and Zhou, X. (2007). "Different responses of MODIS-derived NDVI to root-zone soil moisture in semi-arid and humid regions." *J. Hydrol.*, 340, (1–2), 12–24.
- Weissling, B. P., and Xie, H. (2009). "MODIS biophysical states and Nexrad precipitation in a statistical evaluation of antecedent moisture condition and streamflow." *J. Am. Water Resour. Assoc.*, 45(2), 419–433.



# Assessment of uncertainty in soil test phosphorus using kriging techniques and sequential Gaussian simulation: implications for water quality management in southern Quebec

Alaba Boluwade and C. A. Madramootoo

## ABSTRACT

Missisquoi Bay, located in southern Quebec at the north-eastern extremity of Lake Champlain, is subject to eutrophication arising from excess nutrients, predominantly phosphorus (P), contributed by agricultural watersheds. Factors such as land use pattern, management practices, soil properties and geomorphology have an impact on soil P levels. Land patches under different management practices introduce a cyclic pattern, especially when fitting the variogram. Geostatistics procedures were used to model soil test phosphorus (STP) within the 11 km<sup>2</sup> Castor Watershed in southern Quebec, Canada. An ordinary kriging (OK) method was used to estimate STP at unsampled points, but this had a smoothing effect, resulting in an underestimation of high values and overestimation of low values. Therefore, a more efficient technique was needed to draw predictions from the conditional probability distribution at the simulation grid nodes. A sequential Gaussian simulation (SGS) was adopted for this purpose, and used to create 50 equal probable realizations. Uncertainty was modelled using the E-type (mean) of the realizations, which ranged from 12.5 to 223 mg P kg<sup>-1</sup> soil. The adequate spatial probability pattern for STP is a valuable criterion when seeking to delineate areas of high STP for site-specific best management practices (BMPs).

**Key words** | ordinary kriging, phosphorus, probability map, sequential Gaussian simulation, soil test, uncertainty

## INTRODUCTION

Globally, non-point source pollution (NPS), such as nutrient losses from agricultural watersheds, can lead to the degradation of freshwater resources (Carpenter *et al.* 1998; Hegman *et al.* 1999; Beauchemin & Simard 2000). Freshwater pollution can come from point sources (PS) such as community treatment plants, agricultural dairy effluents or industrial setups such as fertilizer manufacturing plants. Over the years, the government policies on water pollution control have helped to minimize PS pollution. On the other hand, nutrient losses – mostly nitrogen (N) and phosphorus (P) – from agricultural fields are very difficult to control or monitor (Sharpley 2007). N and P losses from agricultural fields that exceed certain pre-defined limits lead to the enrichment of surface waters causing what is termed as eutrophication.

In recent years, the Missisquoi Bay (north-eastern portion of Lake Champlain) in southern Quebec, Canada, has experienced eutrophication due to excess nutrient losses from agricultural watersheds (Hegman *et al.* 1999; Jamieson *et al.* 2005; Deslandes *et al.* 2004; Medalie & Smeltzer 2004; Michaud 2004; Michaud *et al.* 2004, 2005). This has been attributed to the consistent application of fertilizers and organic manures and a ‘lack of prior knowledge’ about the nutrient richness and status of soils in this area over the years (Hegman *et al.* 1999; LCBP 2004; Beaudin *et al.* 2005). This has been reported to have very serious impacts on tourism, recreation and aquatic life in the basin (Priskin 2008).

Differences in soil types and properties generally lead to the differences in the nutrient status across a field, although

Alaba Boluwade (corresponding author)

C. A. Madramootoo

Department of Bioresource Engineering,  
Macdonald Campus of McGill University,  
21111 Lakeshore Road, Sainte-Anne-de-Bellevue,  
Quebec,  
H9X3V9,  
Canada  
E-mail: [alaba.boluwade@mail.mcgill.ca](mailto:alaba.boluwade@mail.mcgill.ca)

other external factors such as agricultural practices, management practices, geomorphology, etc. may also be involved. Crop growth and yield depend heavily on the agricultural field's nutrient status, which may vary widely across the field. This information is very important to the farmers in order to apply the 'exact' quantity of nutrient to meet the needs of the crop, particularly in precision agriculture. This variability occurs in both space and time (Geypens *et al.* 1999), which poses a challenge in finding an accurate covariance function for variogram in spatial modelling. In general, soil N and P are considered to be highly variable compared to other soil properties (Geypens *et al.* 1999; Bennett *et al.* 2004; Grunwald *et al.* 2004).

The land use pattern is another factor that strongly affects the spatial distribution of soil test phosphorus (STP). Arable lands often receive large mineral and organic fertilizer inputs, which make large quantities of P available for loss through erosion or leaching.

Spatial probability mapping and quantification of uncertainty are very important criteria when one wants to delineate areas of high and low STP in order to implement site-specific best management practices (BMPs). Spatial modelling involves both estimation and simulation. Ordinary kriging (OK) is one of the most efficient methods in spatial estimation, while a sequential Gaussian simulation (SGS) can be used for simulation. SGS involves using a Monte Carlo technique to draw predictions from the conditional probability distribution at each of the grid nodes. In other words, SGS involves generating realizations. According to Remy *et al.* (2006), each realization can be obtained by: (i) defining a random path visiting each node of the grid, (ii) obtaining, for each node along the path, a data set for the area in the neighbourhood that includes the observed data set and prior simulated values, (iii) estimating the local conditional cumulative distribution function (ccdf) as a Gaussian distribution with mean given by simple kriging and its variance by the kriging variance, (iv) drawing a value from the Gaussian ccdf and adding the simulated value to the data set, and (v) repeating for another realization.

One of the major disadvantages of the kriging method is that it has a smoothing effect (Goovaerts 1997; Juang *et al.* 2004), resulting in small values being overestimated and large values being underestimated (Goovaerts 1997; Juang *et al.* 2004). The smoothening effect obstructs the

understanding of the variability and detection of patterns of extreme values, such as areas with high STP (Goovaerts 1997; Juang *et al.* 2004). Thus the kriging output only represents a simplistic pattern and does not reproduce spatial detail. This shortcoming can result in a misleading interpretation when the goal of the research is to characterize the areas that have a large amount of STP for site-specific BMPs in the form of management zones. Juang *et al.* (2004) reported that the kriging value at each unsampled location includes an estimation variance, which can cause uncertainty in mapping. Considering this uncertainty is very important, especially when determining 'hot spots' of high STP, which, if linked with P transport pathways, have the potential to cause excess P transport into fresh water.

SGS is a method that is able to determine the multi-point or global uncertainties of STP using the conditional probability distribution of simulated and observed data at the simulation grid nodes (Juang *et al.* 2004; Bivand *et al.* 2008). SGS allows the quantification of uncertainty at several locations simultaneously. This is required in order to assess the reliability of an area in the form of management zones for site-specific BMP applications, as opposed to the estimated output from kriging methods (Goovaerts 1997; Grunwald *et al.* 2004; Juang *et al.* 2004). SGS has been reported to overcome the smoothing limitations of ordinary kriging (Goovaerts 1997; Grunwald *et al.* 2004; Juang *et al.* 2004). SGS takes into consideration not only the spatial variation of observed data at sampled locations, but also the variation in estimation at unsampled locations, which kriging ignores (Grunwald *et al.* 2004; Juang *et al.* 2004). Evaluations of uncertainty for all possible simulation nodes are plausible. Finally, SGS can collect mapping uncertainties at several locations simultaneously, which is 'masked' in ordinary kriging.

In this paper, however, the OK method was used to estimate the STP at unsampled locations. The shortcomings or limitations of OK in terms of the smoothing effect were evaluated. SGS was used to quantify the uncertainties at unsampled locations by generating realizations that have equal probabilities. These realizations honour the conditional data set in each of the simulation nodes. Probability maps above recommended thresholds were developed using SGS and OK. Such a study is very important, especially for the Misisquoi Bay where there are recurring issues of cyanobacteria blooms from the influx of P from agricultural fields.

## MATERIALS AND METHODS

### Study area

Situated near Bedford, QC, the 11 km<sup>2</sup> Castor watershed is the smallest of a number of sub-watersheds in the Pike River Watershed (Figure 1) (Beaudin *et al.* 2005; Deslandes *et al.* 2007). It feeds into the downstream portion of the Pike River, which has its outlet in the Mississquoi Bay, in the north-eastern portion of Lake Champlain. The Castor watershed is in an area of intensive agricultural activity, where land is cropped 44% to corn (*Zea mays* L.), 28% to grass (hay) and 20% to cereals (Beaudin *et al.* 2005). Other agricultural activities include swine, poultry and dairy production. Ranging from 36 to 49 m AMSE, the Castor watershed is relatively flat. This watershed has a total of six soil classes according to the Quebec classification (Michaud 2004), of which the largest proportion (by area) are fine textured (Figure 2). This results in low water

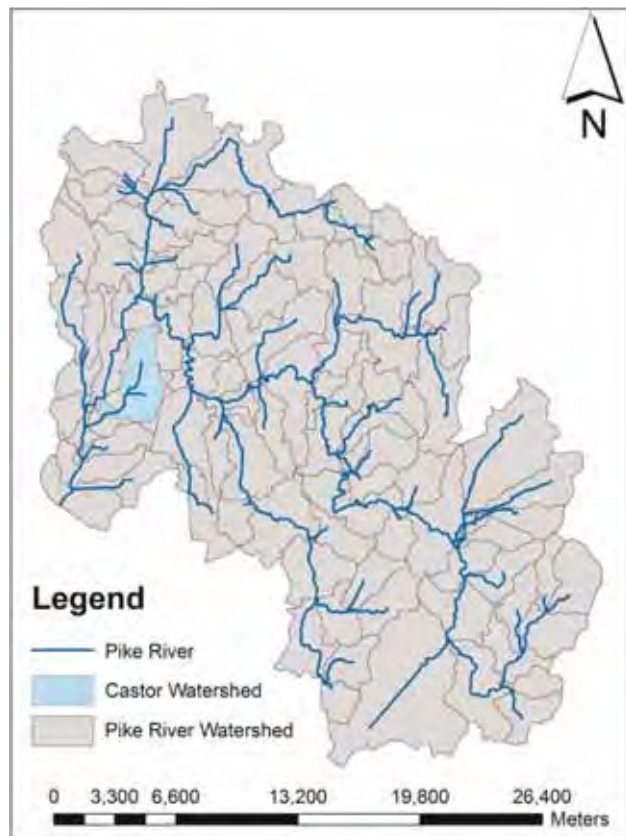


Figure 1 | Geographical location of the study area (Source: LCBP 2004).

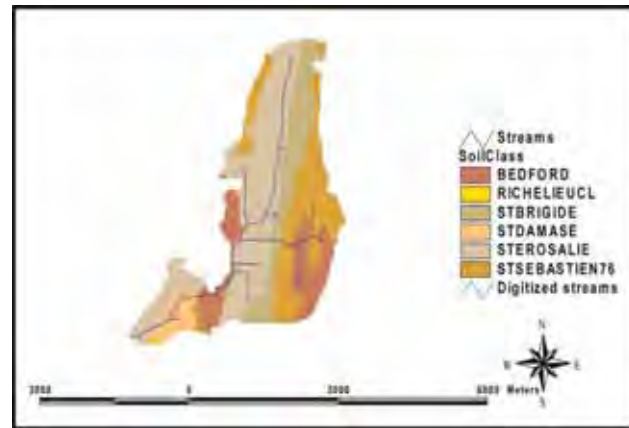


Figure 2 | Soil classes for Castor watershed.

infiltration and poor drainage, which along with wet antecedent moisture conditions and snow melt have contributed to flooding events in recent years. Consequently, sub-surface drains are installed on most farms to remove excess water from the agricultural fields. The water table in this region is reported to be very shallow (Beaudin *et al.* 2005; Deslandes *et al.* 2007). The mean long-term annual precipitation of this study area is about 1,057 mm (MDDEPQ 2003; Michaud 2004; MDDEPQ 2005; MENV 2006).

The major soil series of the watershed are (Cann *et al.* 1946; Cattai 2004; Michaud 2004):

- Ste Rosalie: poorly drained lacustrine and marine clays
- Bedford an Orthic Humic Gleysol, fine-loamy mixed calcareous
- Ste Brigide: an Orthic Humic Gleysol, coarse-loamy mixed calcareous
- St Sebastien: Gleyed Sombric Brunisol, loamy-skeletal, mixed, monoacid, mesic.

### Field sampling technique

A 100 m × 100 m grid (being the smallest that could be used) was imposed over the entire area of the Castor watershed, and each cell had a geo-referenced coordinate at its centre (Figure 3). The study area was divided into three sub-basins (Figure 3) (Beaudin *et al.* 2005). The soil series served as strata for a stratified random sampling, which covered the entire spatial extent of the



**Figure 3** | The geo-referenced sampling points.

study area. Due to limited financial resources and time constraints, it was only feasible to obtain a total of 211 sampling points, geo-referenced and imported into a Garmin Oregon 450 Global Positioning System (GPS). Webster & Oliver (1992) reported that 100–150 sampling points are sufficient to compute a reliable variogram, especially in the area where the spatial variation is isotropic. This (variogram) aids in developing and estimating spatial structure and dependency of the variable of interest and also is a requirement for spatial interpolation using a kriging technique. ArcGIS 10 software (Environmental Systems Research Institute, Redlands, CA) was used to conduct the geostatistical estimation, simulation and visualization. Each of the points was geo-referenced and imported into a GPS, which has a positional accuracy of  $\pm 2$  m. A sampling depth of 0–0.30 m was used in this study. This is the most active layer in terms of P fate and transport. The soil samples were all air-dried, and sieved through a 2 mm size sieve.

### Laboratory analysis

STP was extracted from soil using the Mehlich-3 (Mehlich 1984) extractant. The samples were analysed following the procedure of Sims (2000). An air-dried soil sample (2.5 g) was mixed with 25 mL of Mehlich-3 extractant and shaken for 5 min. at 200 rpm, then filtered through Whatman No. 42 filter paper. The sample filtrates were transferred into the auto-sampler rack of an inductively coupled plasma-optical emission spectrometer (ICP-OES) to assess STP levels. The ICP can analyse about 120 samples for 10 elements within 7 hours.

### Variogram analysis

Matheron (1970) described geostatistics as a field that deals with a random field  $X$  that is linked to a multi-dimensional locus  $u$ , and there exists a relationship between  $X(u)$  and  $X(u+h)$ , where  $h$  is the lag-distance (Ploner & Dutter 2000; Barnes 2011). A variogram, a term used here synonymously with semi-variogram, characterizes the spatial continuity or spatial roughness of a spatial data set (Bivand *et al.* 2008). The variogram plots the semi-variance as a function of distance  $h$ .

One must make a stationary assumption in order to estimate the spatial correlation of the observation data (Bivand *et al.* 2008). A very basic form of assumption of stationarity is that of intrinsic stationarity. This assumes that the process that formulates the samples is a random function  $X(u)$ , composed of a mean and residual (Isaaks & Srivastava 1989; Goovaerts 1997; Bivand *et al.* 2008).

Estimating the spatial correlation from the observation data involves making a stationary assumption (Bivand *et al.* 2008). This assumption states that the process that generated the observation data is a random function  $X(u)$ , which can be expressed as:

$$X(u) = \mu + \epsilon'(u) \quad (1)$$

where:

$X(u)$  is a random function at a location  $u$

$\mu$  is the constant mean and

$\epsilon'(u)$  is the residual or error term.



A constant mean,  $\mu$ , is the expectation of  $X(u)$ :

$$E(X(u)) = \mu \quad (2)$$

Therefore, under this assumption, the variogram,  $\gamma(h)$  can be defined as:

$$\gamma(h) = \left(\frac{1}{2}\right) * E[X(u) - X(u+h)]^2 \quad (3)$$

where  $h$  is the separation distance.

From Equation (3), the variance of  $X$  is constant and the spatial correlation of  $X$  depends not on location  $u$ , but on the separation distance  $h$  (Bivand *et al.* 2008).

If 'isotropy' is assumed in Equation (3),  $h$  can be replaced with  $\|h\|$ . This means that the variogram has directional independence (Bivand *et al.* 2008). Using this isotropic assumption, the variogram from  $N_i$  sample with data pairs  $X(u_i + h)$ ,  $X(u_i - h)$  for a number of distances  $\check{h}_j$  (distance intervals) can be estimated as (Bivand *et al.* 2008):

$$\Gamma(\check{h}_j) = \frac{1}{2N_i} \sum_{i=1}^{N_i} [(X(u_i) - X(u_i + h))^2], \quad \forall h \in \check{h}_j \quad (4)$$

where  $\Gamma$  is the estimate of the sample variogram.

In many geostatistics software products, there are four basic covariance functions (variogram models) along with any positive linear combinations of them that generate the linear model of regionalization (Goovaerts 1997). These four models are (Remy 2004):

(i) Nugget effect model

$$\gamma(h) = \begin{cases} 0 & \text{if } \|h\| = 0 \\ 1 & \text{otherwise} \end{cases} \quad (5)$$

(ii) Spherical model with actual range  $a$

$$\gamma(h) = \begin{cases} \frac{3}{2} \frac{\|h\|}{a} - \frac{1}{2} \left(\frac{\|h\|}{a}\right)^3 & \text{if } \|h\| \leq a \\ 1 & \text{otherwise} \end{cases} \quad (6)$$

(iii) Exponential model with practical range  $a$

$$\gamma(h) = 1 - \exp\left(\frac{-3\|h\|}{a}\right) \quad (7)$$

(iv) Gaussian model with practical range  $a$

$$\gamma(h) = 1 - \exp\left(\frac{-3\|h\|^2}{a^2}\right) \quad (8)$$

These four covariance functions (Equations (5)–(8)) are those most commonly used in geostatistics analysis. These models have a monotonic property (Pyrz & Deutsch 2003). However, there are some situations where the experimental variogram exhibits some cyclic pattern. Presenting a non-monotonic continuity structure, this kind of variogram is often called a hole effect variogram (Pyrz & Deutsch 2003). Neglecting these non-monotonic structures in the analysis may result in misinterpretations in heterogeneity models, such that they do not reproduce the observed patterns of variability (Pyrz & Deutsch 2003).

The mathematical expression that represents a hole effect can be given as (Pyrz & Deutsch 2003):

$$\gamma(h) = c_n \left[ 1 - \cos\left(\frac{h\pi}{a_n}\right) \right] \quad (9)$$

where:

$c_n$  is the partial sill (variance contribution) and  $a_n$  is the range parameter, and other parameters are as defined above.

## Ordinary kriging

Ordinary kriging assumes that the means ( $\mu$ ) are constant in the local neighbourhood of each estimation point rather than the entire domain, that is (Isaaks & Srivastava 1989; Goovaerts 1997; Bohling 2005; Bivand *et al.* 2008):

$$\mu(u_\alpha) = \mu(u) \quad (10)$$

where  $u_\alpha, u$  are location vectors for estimation point and one of the neighbouring data points, indexed by  $\alpha$ .

The kriging estimator can be written as (Bohling 2005):

$$X^*(u) = \mu(u) + \sum_{\alpha=1}^{n(u)} \lambda_\alpha(u) [X(u_\alpha) - \mu(u)] \quad (11)$$



where:

$n(u)$  is the number of data points in the local neighbourhood used for estimation of  $X^*(u)$ .

$\lambda_\alpha(u)$  is the kriging weight assigned to datum  $X(u_\alpha)$  for estimation location  $u$ .

Equation (11) can further be expressed in terms of the expected values of  $X(u)$  and  $X(u_\alpha)$  as Bohling (2005):

$$X^*(u) - m(u) = \sum_{\alpha=1}^{n(u)} \lambda_\alpha(u)[X(u_\alpha) - m(u_\alpha)] \tag{12}$$

where  $m(u), m(u_\alpha)$  are expected values (means) of  $X(u)$  and  $X(u_\alpha)$ , respectively.

### Sequential Gaussian simulation

The procedure for generating a sequential Gaussian simulation consists in defining a regularly spaced grid that covers the entire watershed area and thereby establishes a random path for the entire grid. The process follows the steps enumerated below (Chilès & Delfin 1999; Bourennane et al. 2007; Bivand et al. 2008; Delbari et al. 2009).

(a) *Transform variable*: This constitutes a statistical transformation of the variable of interest to normality using a multi-Gaussianity assumption. It is possible to transform the marginal distribution of the variable into a Gaussian distribution and simulation is performed on the transformed variable (Remy 2004). The transformation of the variable  $X(u)$  with the cumulative distribution function cdf  $Fz$  into a standard normal variable  $Y(u)$  with the cdf  $G$  is given as (Remy 2004):

$$Y(u) = \frac{F[X(u)]}{G} \tag{13}$$

- (b) *Estimate covariance*: Calculation of the covariance of the transformed values in (a).
- (c) *Grid path analysis*: Definition of a grid path so as to visit all the nodes to be simulated and so that no empty node is visited along the path.
- (d) *Estimate using Simple Kriging (SK)*: Estimation of simple kriging mean and variance conditioned by the available data and prior predicted values.

- (e) *Apply Monte Carlo technique*: The Monte Carlo technique will be used to draw a value from the conditional Gaussian distribution as defined by the SK mean and variance.
- (f) *Add simulated value*: Add the simulated value to the conditioning data set.
- (g) *Iterate all nodes*: Return to (c) until all nodes are visited.

*Back transform the simulated values*: Transformation to the original values to obtain a realization (Remy 2004), e.g. transforming the simulated values  $y_1, y_2, \dots, y_N$  into  $x_1, x_2 \dots x_N$  such that:

$$x_i = F^{-1}(G(y_i)) \quad i = 1, 2, \dots, N \tag{14}$$

## RESULTS AND DISCUSSION

### STP data analysis

Descriptive statistics of STP levels across the entire the Castor watershed are shown in Table 1. Over the full watershed there was a wide range of STP values (4.29 to 306.55 mg kg<sup>-1</sup>). As shown in Figure 4, where the whiskers indicate the 10th and 90th percentiles of measured STP values, these varied according to land use. Pastures showed

**Table 1** | Descriptive statistics for soil test phosphorus (mg P kg<sup>-1</sup> soil) across the full Beaver watershed, and by grouped land uses

Statistic	Land use categories			
	Full watershed	Pasture	Arable (corn and soy)	Low-P (residential, tall grass, forest)
Mean	71.0	91	76	49
Standard error	3.9	15.8	4.8	7.3
Median	55	69	58	37
Standard deviation	57	61	57	53
Sample variance	3356	3762.4	3299.9	2830.1
Kurtosis	3.70	-0.59	3.37	12.14
Skewness	1.86	0.88	1.77	3.21
Range	302	189.5	301.2	298.6
Minimum	4.29	21	5.2	4.2
Maximum	306.55	211.54	306.55	302.97

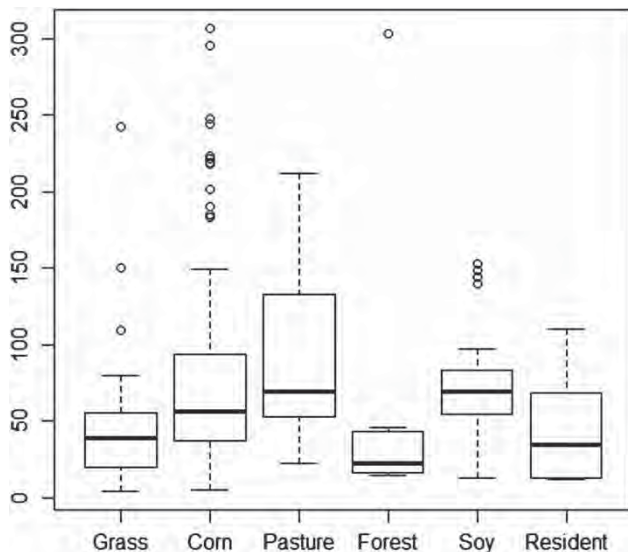


Figure 4 | Soil test phosphorus ( $\text{mg kg}^{-1}$ ) by land use pattern in the Castor watershed.

the highest STP levels, while forested areas showed the lowest. This makes sense since forests generally do not receive P inputs as either manure or fertilizer.

Kriging is based on the stationarity assumption for the mean value. Given the wide variability in STP levels it would be a challenge to develop an efficient interpolated surface for these unstructured and noisy data using this method; however, these data clearly showed a distinct and less variable mean value according to the land use pattern. Accordingly, for the kriging assumption to be valid for this study, the data were divided into three land use categories, based on their estimated STP values. Descriptive statistics for these categories are presented in Table 1:

- Pasture fields
- Arable lands: corn and soy fields
- Low P land use areas: residential, grasslands (prairie tall grasses) and forested areas.

The highest STP values were found in areas falling into the arable lands category. Therefore using this land use class as a representative of the study will not mask high STP areas. Also, there are similarities between the descriptive statistics of STP for the lumped land use classes and the arable lands (Table 1). This will allow us to assume stationarity of OK. Therefore, the data used in this research were reduced to arable lands only.

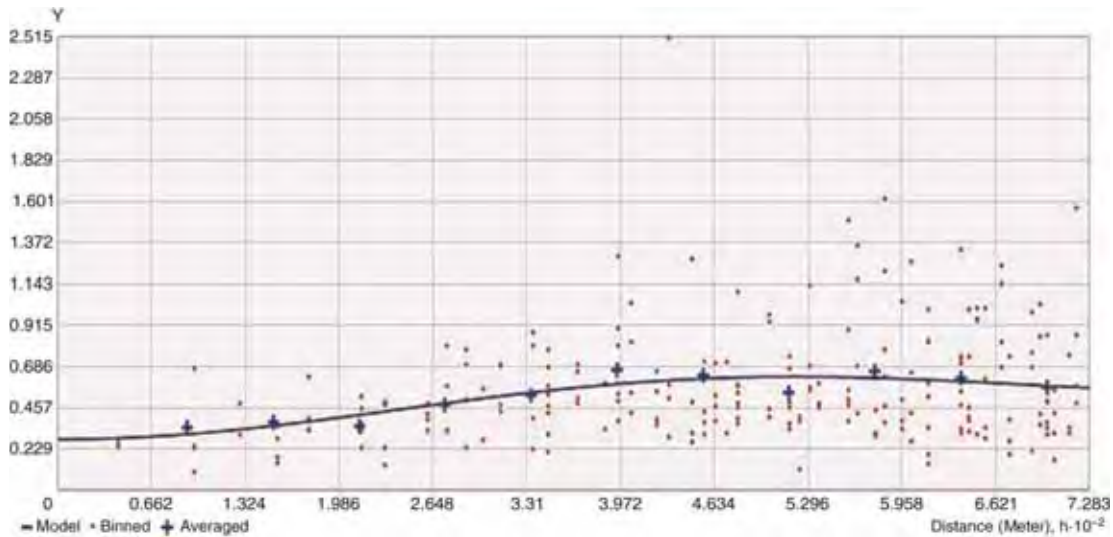
A one-way analysis of variance (ANOVA) (unequal sample sizes) was applied for the groups since the annual phosphorus application and its effects were considered as the only factor across the categories. The mean values for the pastures, arable lands and low-P land categories were 91, 77 and 49  $\text{mg kg}^{-1}$ , respectively. There was evaluation done on whether differences between these means were statistically significant or simply based on random sampling error. The test result showed an F-value of 5.673, while the critical F was 3.039 ( $\alpha = 0.05$ ). This resulted in a P-value of 0.00399. The null hypothesis (equal means) was rejected, indicating that the difference between these groups was not due to random sampling error. Therefore, it would not be correct to combine and model the three groups together. In fact, it would be a challenge to get a good covariance function fit for such combinations. Arable lands (corn and soy) were therefore chosen for spatial modelling.

### Variogram and spatial modelling

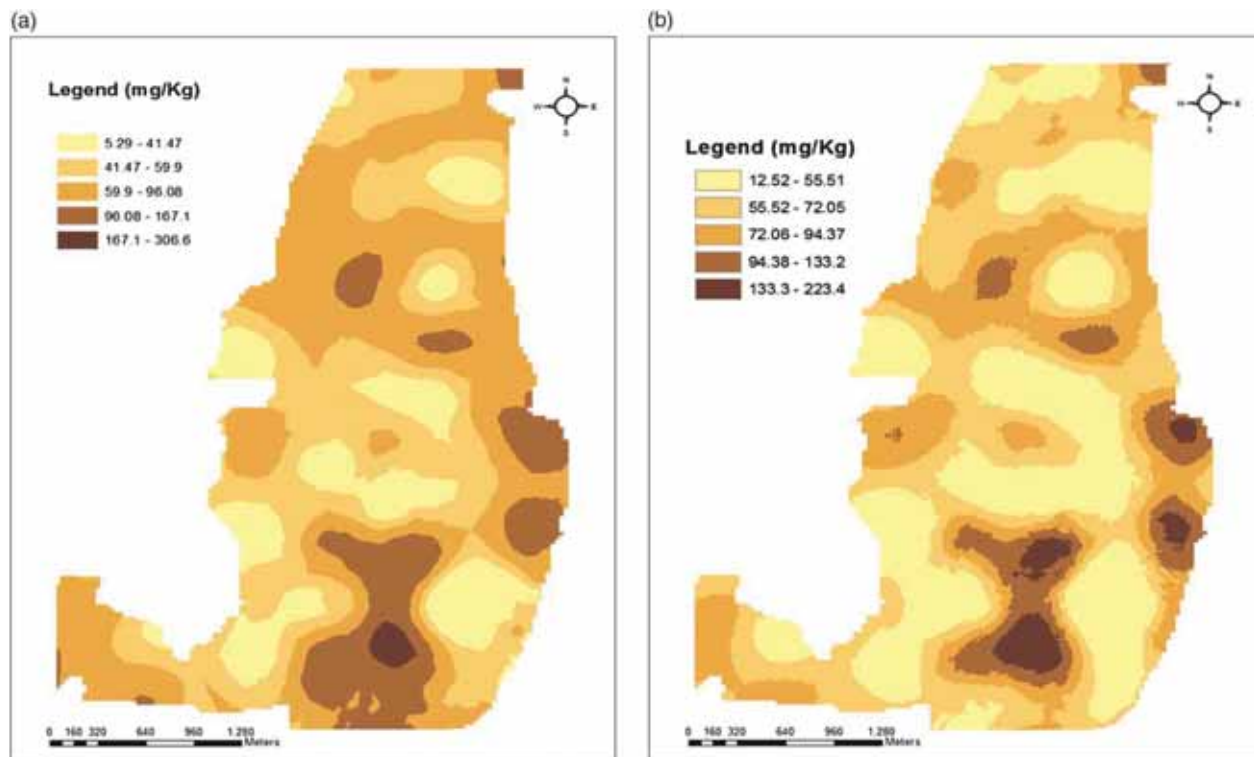
Expert knowledge (through the authors' experience in geostatistics estimation of soil attributes) was used to generate the covariance function for the experimental variogram (Figure 5). There is evidence of a hole effect in the experimental variogram. Therefore a hole effect covariance function was fitted to the data.

The spatial patterns of STP obtained using the OK and STP (E-type, i.e. mean of realizations) are shown in Figures 6(a) and 6(b), respectively. The higher values were close to the areas associated with intensive agriculture and animal husbandry. The STP map generated by OK can be seen to be much smoother compared with that derived from the SGS realization technique. SGS method generated 50 realizations where a random path was defined and the algorithm visited each node of the grids using the measured values as conditioning data. In terms of model performance, the OK method resulted in coefficients of correlation and determination of 0.85 and 0.73, respectively (Figure 7).

The spatial uncertainty obtained using SGS is shown for two realizations (Figures 8(a) and 8(b)). Figures 9(a) and 9(b) capture the best and worst scenarios (minimum and maximum) developed using SGS. Stochastic simulation using SGS generated multiple equiprobable realizations of STP compared to the mean estimation using OK (Bohling



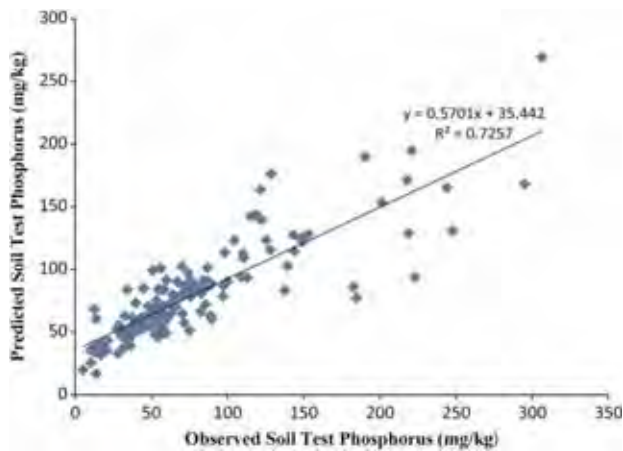
**Figure 5** | The local (regional) variogram of soil test phosphorus collected in the Castor watershed. (The semivariance is positive definite and normalized by logarithmic transformation with nugget = 0.2831, partial sill = 0.2856 and range = 728 m. ArcGIS 10 did an automatic back-transformation of results.)



**Figure 6** | Soil test phosphorus estimated by (a) ordinary kriging and (b) sequential Gaussian simulation (E-type).

2005). In essence, SGS technique added some noise to the simulation in order to overcome the smoothing effect using OK (Bohling 2005). From the foregoing, SGS gives a better representation of the spatial variability of STP and

provides a means for quantifying the uncertainty. Therefore, coupled with the E-type map, these SGS maps provide more realistic estimates of actual conditions than a single output map using the OK technique (Grunwald *et al.* 2004).



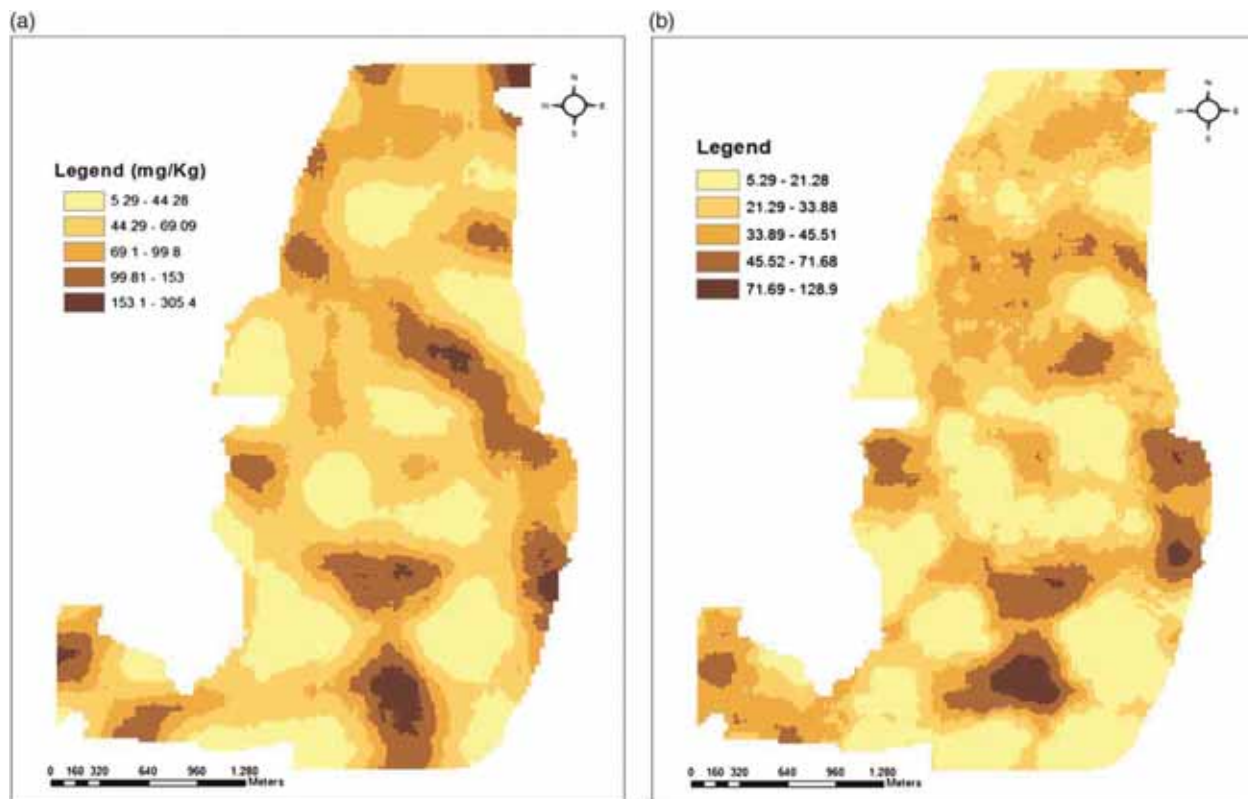
**Figure 7** | Cross-validation model performance using the hole effect covariance function in ordinary kriging (OK).

A threshold of  $60 \text{ mg kg}^{-1}$  STP was recommended by Sharpley *et al.* (2003) as being the necessary level of P in the soil for plant uptake and sustainable growth. Above this value, there is a propensity for P to be washed into fresh water during runoff or erosion (if the land is exposed).

This threshold value was used to develop probability maps using both OK and SGS methods (Figure 10). SGS estimated lower probabilities in the risk class [0.8 to 1.0] with 18% of all pixels above threshold, while OK estimated probabilities in the risk class [0.8 to 1.0] in 14% of all pixels. This can be attributed to OK having overestimated low values or underestimated high values (Isaaks & Srivastava 1989; Webster & Oliver 2001; Grunwald *et al.* 2004). Probabilities of P loss exceeding 0.50 accounted for 48 and 49% of all pixels when using the SGS and OK methods, respectively. Some areas classified as exceeding the threshold in OK were below the threshold in SGS. Clearly, OK and SGS methods arrived at different results from the same data set.

### Implications for water quality and site-specific nutrient management in Quebec

These results have some impact on water quality management, especially for Quebec. The usual method of



**Figure 8** | A subset of realizations using sequential Gaussian simulation, (a) Realization #1, (b) Realization #2.



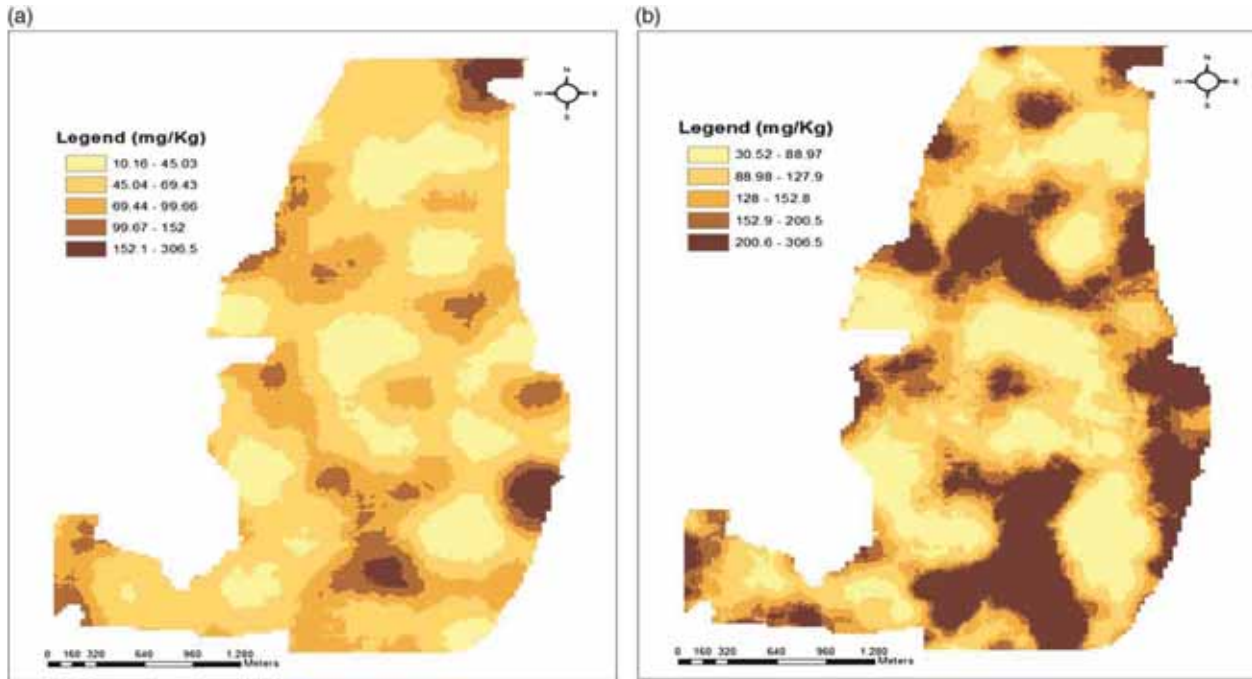


Figure 9 | Scenarios for soil test phosphorus developed using sequential Gaussian simulation, (a) best (minimum) scenario, (b) worst (maximum) scenario.

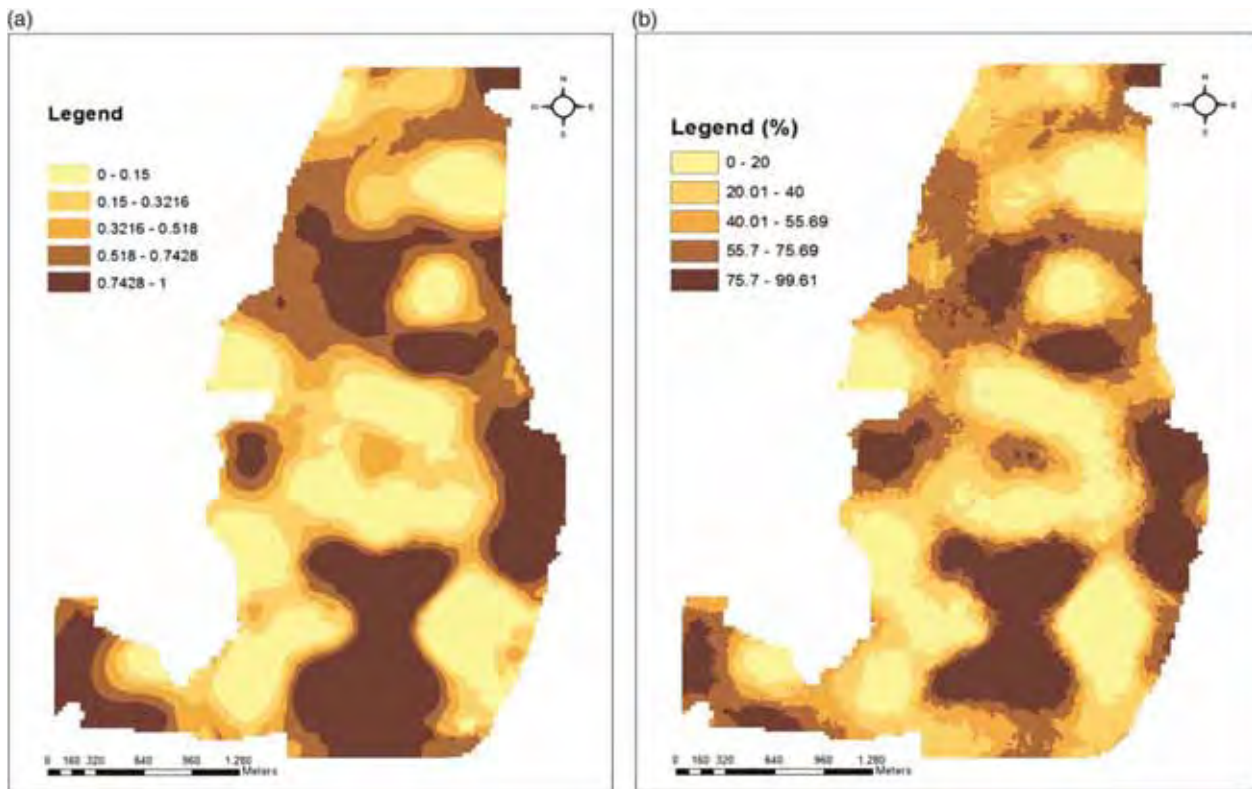


Figure 10 | Probability of exceeding the soil test phosphorus threshold of 60 mg kg<sup>-1</sup> determined using (a) ordinary kriging and (b) sequential Gaussian simulation.



mapping soil attributes (kriging) has inherent uncertainty and a smoothing effect, such that decision-making using this method may be misleading. Low and high values of STP may be masked. The arable lands in the Beaver watershed have a mean STP of  $70 \text{ mg kg}^{-1}$ , which exceeds the recommended threshold of  $60 \text{ mg kg}^{-1}$  necessary for plant growth. The excess  $10 \text{ mg kg}^{-1}$  has the potential of being washed into fresh water systems. From a lake management perspective, there is a standing agreement between the province of Quebec and the State of Vermont for a target reduction of  $97.2 \text{ Mg yr}^{-1}$  in the total P load entering Missisquoi Bay. The reduction goal for Vermont is  $58.3 \text{ Mg yr}^{-1}$  (60%), while that for Quebec is  $38.9 \text{ Mg yr}^{-1}$  (40%) (Vermont, New York and Quebec, 1996; [Medalie & Smeltzer 2004](#)). Quebec has yet to meet this target, as recurrent algal blooms in the bay support this claim ([MDDEPQ 2009](#); [CBS 2013](#)). The P and sediment fluxes are coming from intensive agricultural fields (including this study site).

Therefore, there is a need to intensify the implementation of various BMPs that can reduce nutrient flux into fresh water. Some of these include conservation tillage and buffer strips, as well as crop nutrient management. Farmers in these areas should be encouraged to apply only the right amount of fertilizer and manures. Excessive applications of these build up in the soil and have the potential to flow into surface and ground water. Also, conservation buffer strips have the ability to slow down runoff and trap nutrient and sediment from entering fresh water. The riparian vegetation of at least 3 m ([Bradley et al. 2011](#); [LCBP 2012](#)) from the stream banks has the potential to reduce the movement of nutrients and sediments into streams and lakes.

Overall, this research provides a salient guide for site-specific BMPs for high STP areas identified in [Figures 8\(a\)](#) and [8\(b\)](#), using the aforementioned BMPs, recommendations and much more.

Overall, the uncertainty quantification of highly variable soil parameters like STP requires careful statistical consideration. Whenever the desire is to classify the watershed into areas greater than some pre-defined limit in terms of a threshold STP concentration, SGS gives more realistic maps and scenarios instead of a single map derived in OK.

The major advantage of the SGS technique over methods such as OK or a Thiessen Polygon (TP) approach is that SGS provides an estimate of both the mean and the standard

deviation of the variable at each grid node ([Lin 2008](#)). This implies that the variable can be represented at each grid node as a random variable following a Gaussian distribution ([Lin 2008](#)). Instead of using the mean as the estimate at each node, SGS chooses a random deviate (through a Monte Carlo technique) from this Gaussian distribution, selected according to a uniform random number representing the probability level ([Lin 2008](#)). SGS generates a set of realizations that have statistics similar to that of the conditioning data set and equally quantify the spatial uncertainty ([Lin 2008](#)).

Another important aspect of this paper is its relevance in the areas of nutrient status of this watershed. Soil types and properties generally induce the differences in the nutrient status across the field, although there are other external factors such as agricultural practices. Recently, there has been great interest in site-specific agriculture ([Geypens et al. 1999](#); [Adamchuk et al. 2002](#)). Plant growth and yield depend on the nutrient status of the field. This status varies across the field. The variability also exists in space and time. Generally P is considered to be highly variable compared to other properties ([Geypens et al. 1999](#); [Bennett et al. 2004](#); [Grunwald et al. 2004](#)). Farmers' benefits lie on the ability to apply a variable rate of fertilizer or manure. This is only possible if the experts are able to give the correct site-specific information and recommendations ([Geypens et al. 1999](#)). This paper has been able to quantify precise information about the nutrient status of the watershed through geostatistical techniques, and identify vulnerable areas where the risk of surface runoff when combined with the presence of excess P has the potential to erode P into fresh water. This risk is high in areas where STP exceeds the recommended threshold of what plant really needed.

---

## CONCLUSION

Many soil property maps are based on OK. This technique masks areas of high or low values because of its smoothing effect. Although kriging techniques are still used in digital soil mapping, their results can be misleading because of this smoothing effect (underestimation of high values and overestimation of low values). SGS generates multiple realizations,

including an error component, which are absent in classical interpolation techniques such as OK (Lin 2008). This study demonstrates the application of OK and sequential SGS. The OK technique is based on the procedure of kriging interpolation, which provides the best linear unbiased prediction of a value in terms of minimum error variance (Chilès & Delfiner 1999). As a consequence of management practices, high STP levels were evident in pasture lands. A high coefficient of variation illustrates the lack of spatial pattern or structure in the STP. When the watershed is divided into arable and non-arable lands, there is evidence of an undamped non-monotonic covariance function. This shows that in spatial modelling of an environmental variable that is impacted by management practices, it should be divided into those factors that have influence upon it.

Non-monotonic variogram structures, or hole effects, provide a better variogram fit. The hole effect helps in capturing the impact of the management practices based on the land-use pattern. This form of hole effect is often neglected in geostatistical modelling, but its usage in this paper provides valuable information concerning spatial uncertainty. Neglecting this important information in variogram modelling often gives misleading model results. The OK method provides an interpolated surface that captures some of the hotspots, but the SGS method provides a better result using a Monte Carlo technique to draw from the conditional probability distribution of the nodes of the simulation grids. The realizations from SGS have equal probabilities. The realizations represent a spatial pattern without the smoothing effect of OK. Several arable land parcels exceeded the STP threshold of 60 mg kg<sup>-1</sup> needed for plant growth. Therefore, for Quebec to meet the targeted reduction of 38.9 Mg yr<sup>-1</sup> in total P load entering the Missisquoi Bay there is more work to do in terms of manure and fertilizer applications. In fact, as of 17 August 2012 (reported by NIWR 2012), blue-green algae had caused the death of thousands of fish in Missisquoi Bay. The excess P that caused the bay's nutrient enrichment came from intensive agricultural fields (including this study site).

From the foregoing, in terms of management decision, SGS provides several scenarios and situations with several maps instead of only one map as obtained with the OK method. Thus, for site-specific BMP applications to reduce P flow into fresh water, the SGS technique, which takes

into account uncertainties, gives a more reliable result. This information is very important for deciding where to target BMPs. The greatest utility of this research is that agronomists can use the maps and information developed to target sites and various P reduction measures to specific areas where they will achieve the most benefit.

## ACKNOWLEDGEMENTS

This study was funded by the Brace Water Management Center, McGill University, MacDonald Campus, Sainte-Anne-de-Bellevue, Quebec, Canada. Helen Lelande and Dr Kebba Sabally helped in the laboratory analysis of the soil samples. All the summer students who participated in the field and laboratory analysis of the soil samples are much appreciated.

## REFERENCES

- Adamchuk, V., Morgan, M., Dobermann, A. & Brouder, S. 2002 Feasibility of On-the-go Mapping of Soil Nitrate and Potassium Using Ion-Selective. University of Nebraska-Lincoln. Paper No. 02-1183. Available at: [http://bse.unl.edu/adamchuk/presentations/ASAE\\_2002.pdf](http://bse.unl.edu/adamchuk/presentations/ASAE_2002.pdf).
- Barnes, R. 2011 Variogram Tutorial Golden Software, Inc. <http://www.goldensoftware.com/variogramTutorial.pdf>.
- Beauchemin, S. & Simard, R. R. 2000 Phosphorus status of intensively cropped soils of the St. Lawrence lowlands. *Proc. Soil Sci. Soc. Am.* **64**, 659–670.
- Beaudin, I., Deslandes, J., Michaud, A. & Desjardins, J. 2005 Using SWAT for Water, Sediment and Nutrient Transport in Agricultural Areas: A Case Study in Southern Quebec. Available at: [www.aic.ca/conferences/pdf/2005/Isabelle\\_Beaudin\\_ENG.pdf](http://www.aic.ca/conferences/pdf/2005/Isabelle_Beaudin_ENG.pdf).
- Bennett, E. M., Carpenter, S. R. & Clayton, M. 2004 Soil Phosphorus Variability: Scale-Dependency in an Urbanizing Agricultural Landscape. *Landscape Ecology: Landscape Ecology* **20** (4), 389–400. Available at: [http://nrs-staff.mcgill.ca/bennett/pdfs/04\\_Bennett\\_LandscapeEcology.pdf](http://nrs-staff.mcgill.ca/bennett/pdfs/04_Bennett_LandscapeEcology.pdf). Accessed August 2011.
- Bivand, R. S., Pebesma, E. J. & Gómez-Rubio, V. 2008 *Applied Spatial Data Analysis with R*. Springer, New York, NY.
- Bohling, G. 2005 KRIGING. Kansas Geological Survey. Available at: <http://people.ku.edu/~gbohling/cpe940/Kriging.pdf>. Accessed 26 April, 2013.
- Bourennane, H., King, D., Couturier, A., Nicoulaud, B., Mary, B. & Richard, G. 2007 Uncertainty assessment of soil water content spatial patterns using geostatistical simulations: an empirical comparison of a simulation accounting for single attribute and

- a simulation accounting for secondary information. *Eco. Model.* **205**, 323–335.
- Bradley, R. L., Whalen, J., Chagnon, P.-L., Lanoix, M. & Alves, M. C. 2011 Nitrous oxide production and potential denitrification in soils from riparian buffer strips: Influence of earthworms and plant litter. *Appl. Soil Ecol.* **47**, 6–13.
- Canadian Biodiversity Strategy (CBS) 2013 Nutrient Loading and Algal Blooms. Available at: <http://www.biodivcanada.ca/default.asp?lang=En&n=BAE73048-&offset=2&toc=show>. Accessed 16 May, 2013.
- Cann, D. B., Lajoie, P. & Stobbe, P. C. 1946 *Soil Survey of Shefford, Brome and Missisquoi Counties in the Province of Quebec*. Publication #3M-10358-12-47. King's Printer and Controller of Stationery, Ottawa, ON. Available at: <http://sis.agr.gc.ca/cansis/publications/pq/pq11/intro.html>.
- Carpenter, S. R., Caraco, N. F., Correl, D. L., Howath, R. W., Sharpley, A. N. & Smith, V. H. 1998 Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecol. Appl.* **8** (3), 559–486.
- Cattai, J. 2004 La détection de contours dans l'approche stratifiée par champs pour la classification d'un milieu agricole à l'aide de données LANDSATM, Cas du bassin versant de la Rivière aux Brochets, sud du Québec (*Contour detection in a stratified field-scale approach to the classification of an agricultural landscape from LANDSATM data*, case study of the Pike River watershed, southern Quebec). MSc thesis, Département de géographie et télédétection, Faculté des lettres et sciences humaines, Université de Sherbrooke, Sherbrooke, QC.
- Chilès, J. P. & Delfiner, P. 1999 *Geostatistics: Modeling Spatial Uncertainty, Second Edition*. Wiley, NY, USA.
- Delbari, M., Afrasiab, P. & Loiskandl, W. 2009 Using sequential Gaussian simulation to assess the field-scale spatial uncertainty of soil water content. *Catena* **79**, 163–169.
- Deslandes, J., Beaudin, I., Michaud, A., Bonn, F. & Madramootoo, C. A. 2007 Influence of landscape and cropping system on phosphorus mobility within the Pike River watershed of Southwestern Quebec: model parameterization and validation. *Can. Water Resour. J.* **32** (1), 21–42.
- Deslandes, J., Michaud, A. & Bonn, F. 2004 Use of GIS and remote sensing to develop indicators of phosphorus non-point source pollution in the Pike River basin. In: *Lake Champlain: Partnerships and Research in the New Millennium* (T. O. Manley, P. L. Manley & T. B. Mihuc, eds). Kluwer Academic/Plenum Pub., New York, NY, USA, pp. 271–290.
- Geypens, M., Vanongeval, L., Vogels, N. & Meykens, J. 1999 Spatial variability of agricultural fertility parameters in a gleyic podzol of Belgium. *Precis. Agr.* **1**, 319–326.
- Goovaerts, P. 1997 *Geostatistics for Natural Resources*. Oxford University Press, New York, NY, USA.
- Grunwald, S., Reddy, K. R., Newman, S. & DeBusk, W. F. 2004 Spatial variability, distribution and uncertainty assessment of soil phosphorus in a South Florida wetland. *Econometrics* **15**, 811–825.
- Hegman, W., Wang, D. & Borer, C. 1999 *Estimation of Lake Champlain Basinwide Nonpoint Source Phosphorus Export*. Lake Champlain Basin Program Technical Report No. 31. LCBP, Grand Isle, Vermont.
- Isaaks, E. H. & Srivastava, R. M. 1989 *An Introduction to Applied Geostatistics*. Oxford University Press, New York, NY.
- Jamieson, A., Madramootoo, C. A. & Enright, P. 2003 Phosphorus losses in surface and subsurface runoff from a snowmelt event on an agricultural field in Quebec. *Can. Biosystems Eng.* **45** (1), 1–7.
- Juang, K., Chen, Y. & Lee, D. 2004 Using sequential indicator simulation to assess the uncertainty of delineating heavy-metal contaminated soils. *Environ. Pollut.* **127**, 229–238.
- Lake Champlain Basin Program (LCBP) 2004 *Nature of the Basin. Lake Champlain Basin Atlas*. LCBP, Grand Isle, Vermont. Available at: [http://www.lcbp.org/ATLAS/HTML/nat\\_region.htm](http://www.lcbp.org/ATLAS/HTML/nat_region.htm).
- Lake Champlain Basin Program (LCBP) 2012 Chapter 4: Reducing Phosphorus Pollution. Available at: <http://plan.lcbp.org/ofa-database/chapters/reducing-phosphorus-pollution/increase-riparian-buffers-minimum-of-3m-in-l>. Accessed 22 April, 2013.
- Lin, Y.P. 2008 Simulating spatial distributions, variability and uncertainty of soil arsenic by geostatistical simulations in geographic information systems. *Open Environ. Sci.* **2**, 26–33.
- Matheron, O. 1970 *The Theory of Regionalized Variables and Its Application*. Les Cahiers du CGMM, Fontainebleau.
- Medalie, L. & Smeltzer, E. 2004 Status and trends of phosphorus in Lake Champlain and its tributaries, 1990–2000. In: *Lake Champlain: Partnerships and Research in the New Millennium* (T. O. Manley, P. L. Manley & T. B. Mihuc, eds). Kluwer Academic/Plenum Pub., New York, NY, USA, pp. 191–219.
- Mehlich, A. 1984 Mehlich 3 soil test extractant: a modification of Mehlich 2 extractant. *Comm. Soil Sci. Plant Analysis* **15**, 1409–1416.
- Michaud, A. R. 2004 *Indicateurs Agroenvironnementaux Adaptes A la Gestion De Project Cibles Sur la Prévention De la Pollution Diffuse par le Phosphore*. Faculté Des Sciences De l'agriculture et De l'alimentation. Université Lava, Québec.
- Michaud, A. R., Lauzier, R. & Laverdière, M. R. 2004 Temporal and spatial variability in nonpoint source phosphorus in relation to agricultural production and terrestrial indicators: the Beaver Brook case study. In: *Lake Champlain: Partnerships and Research in the New Millennium* (T. O. Manley, P. L. Manley & T. B. Mihuc, eds). Kluwer Academic/Plenum Pub., New York, NY, USA, pp. 97–121.
- Michaud, A. R., Lauzier, R. & Laverdière, M. R. 2005 Mobilité du phosphore et intervention agroenvironnementale en bassin versant agricole: Étude de cas du Ruisseau au Castor, tributaire de la Rivière aux Brochets, Québec (Phosphorus mobility and agroenvironmental intervention on an agricultural watershed: case study of Beaver Brook, tributary of the Pike River, Québec). *Agrosol* **16** (1), 47–59.
- Ministère de l'environnement du-Quebec (MENV) 2006 Météorologique data. Farnham (7022320) and Philipsburg (7026040) Stations. Direction du suivi de l'état de

- l'environnement. Suive de l'information sur le milieu atmospherique (SIMAT).
- Ministère du Développement durable, de l'Environnement et des Parcs Quebec (MDDEPQ) 2009 Bilan des lacs et cours d'eau touchés par une fleur d'eau d'algues bleu-vert au Québec. De 2004 à 2009 [online]. [http://www.mddep.gouv.qc.ca/eau/algues-bv/bilan/liste\\_comparative.asp](http://www.mddep.gouv.qc.ca/eau/algues-bv/bilan/liste_comparative.asp). Accessed 16 May, 2013.
- Ministère du Développement durable, Environnement et Parcs du Quebec (MDDEPQ) 2003 Base de données Climatologique du Quebec de 1997 à 2003 (Climatological database for Quebec, 1997–2003). Service de l'information sur le milieu atmosphérique, Direction du suivi de l'état de l'environnement. Available at: [http://www.hc-sc.gc.ca/ewh-semt/pubs/eval/inventory-repertoire/climatologie\\_f.html](http://www.hc-sc.gc.ca/ewh-semt/pubs/eval/inventory-repertoire/climatologie_f.html).
- Ministère du Développement durable, Environnement et Parcs du Quebec (MDDEPQ) 2005 Le réseau hydrométrique Québécois 1998 à 2003 (Quebec's hydrological network 1998–2003). MDDEPQ, Centre d'expertise hydrique du Quebec. Available at: <http://www.cehq.gouv.qc.ca/hydrometrie/reseau/index.htm>.
- National Institute for Water Resources (NIWR) 2012 *Blue-green Algae Kills Thousands of Fish in Missisquoi Bay*. Vermont Water Resources & Lake Studies Center. Available at: <http://www.uvm.edu/rsenr/vtwater/?Page=news&storyID=14222&category=vwrlsc>. Accessed 22 April, 2013.
- Ploner, A. & Dutter, R. 2000 New directions in geostatistics. *J. Stat. Plann. Infer.* **91**, 499–509.
- Priskin, J. 2008 Implications of Eutrophication for Lake Tourism in Québec, Téoros [En ligne], 27-2 2008, mis en ligne le 01 juin 2009, Consulté le 27 Octobre 2011. Available at: <http://teoros.revues.org/139>. Accessed August 2011.
- Pyrzcz, M. J. & Deutsch, C.V. 2003 *The Whole Story on the Hole Effect*. Centre for Computational Geostatistics. Department of Civil and Environmental Engineering. University of Alberta, Edmonton, Alberta, Canada. Available at: [http://www.gaa.org.au/pdf/gaa\\_pyrzcz\\_deutsch.pdf](http://www.gaa.org.au/pdf/gaa_pyrzcz_deutsch.pdf). Accessed May 17, 2012.
- Remy, N. 2004 Geostatistical Earth Modeling Software: User's Manual. Available at: [http://sgems.sourceforge.net/old/doc/sgems\\_manual.pdf](http://sgems.sourceforge.net/old/doc/sgems_manual.pdf). Accessed May 17, 2012.
- Remy, N., Boucher, A. & Wu, A. 2006 SGeMS User's Guide. Available at: <http://www.ebah.com.br/content/ABAAAA-Tm4AF/manual-sgems>. Accessed February 20, 2013.
- Sharpley, A. N. 2007 Modeling phosphorus movement from agriculture to surface waters. In *Modeling Phosphorus in the Environment: State of the Art* (D. Radcliffe & M. Cabrera, eds). CRC Press, Boca Raton, FL, USA, pp. 3–19.
- Sharpley, A. N., Daniel, T., Sims, T., Lemunyon, J., Stevens, R. & Parry, R. 2003 *Agricultural Phosphorus and Eutrophication, 2nd edition*. US Department of Agricultural Research Service, ARS-149, 44 pp.
- Sims, J. T. 2000 Soil test P: Mehlich-3. In: *Methods of Phosphorus Analysis for Soils, Sediments, Residuals, and Waters* (G. M. Peirzynski, ed.). Southern Cooperative Series Bulletin No. # 396. [http://www.sera17.ext.vt.edu/Documents/Methods\\_of\\_P\\_Analysis\\_2000.pdf](http://www.sera17.ext.vt.edu/Documents/Methods_of_P_Analysis_2000.pdf).
- Webster, R. & Oliver, M. A. 1992 Sample adequately to estimate variograms of soil properties. *J. Soil Sci.* **43**, 177–192.
- Webster, R. & Oliver, M. A. 2001 *Geostatistics for Environmental Scientists*. J. Wiley & Sons, Chichester, UK.



## DROUGHT STRESS

**Legume Production and Irrigation Strategies in the Aral Sea Basin: Yield, Yield Components, Water Relations and Crop Development of Common Bean (*Phaseolus vulgaris* L.) and Mungbean (*Vigna radiata* (L.) Wilczek)**

M. Bourgault<sup>1,3</sup>, C. A. Madramootoo<sup>1</sup>, H. A. Webber<sup>1</sup>, P. Dutilleul<sup>1</sup>, G. Stulina<sup>2</sup>, M. G. Horst<sup>2</sup> & D. L. Smith<sup>1</sup>

1 Plant Science Department, Faculty of Agricultural and Environmental Sciences, McGill University Ste-Anne-de-Bellevue, QC, Canada

2 Central Asian Research Institute of Irrigation (SANIIRI) Tashkent, Republic of Uzbekistan

3 CSIRO Plant Industry, St Lucia, Qld, Australia

**Keywords**

alternate furrow irrigation; *Phaseolus vulgaris*; regulated deficit irrigation; Uzbekistan; *Vigna radiata*; water relations

**Correspondence**

D. L. Smith  
Department of Plant Science  
McGill University, Macdonald Campus  
21,111 Lakeshore Road  
Ste-Anne-de-Bellevue, QC  
Canada  
Tel.: +1 514 398 7866  
Fax: +1 514 398 7897  
Email: donald.smith@mcgill.ca

Accepted November 16, 2012

doi:10.1111/jac.12016

**Abstract**

With world population expected to reach 9.2 billion people by 2050, improved irrigation methods will be needed to increase the productivity of agricultural land and improve food supply worldwide. The objective of this work was to examine the effect of regulated deficit irrigation (RDI) and alternate furrow irrigation (AFI) on the yield and yield components of two legume species (common bean and mungbean) produced as a second crop following winter wheat in Uzbekistan, Central Asia. Water relations and crop development were also examined. The research was conducted during two successive growing seasons in the Fergana valley. Production of mungbean using the severe stress RDI treatment in combination with AFI resulted in the highest yields with the lowest quantity of applied water in 2004. In addition, yields of common bean in the moderate stress treatment were not different from the recommended schedule, although irrigation events were decreased from 4 to 2. AFI did not reduce yields, and it did not interact with RDI to reduce yields further. In general, mungbean yields were higher than those of common bean. The combination of AFI and RDI can allow legume production with reduced water inputs.

**Introduction**

Irrigated agricultural systems represent approximately 17 % of the cultivated area and contribute 40 % of total crop production worldwide (Kijne et al. 2003). Increased use of supplemental irrigation water in rainfed agriculture and improved irrigation methods will be needed to increase the productivity of agricultural land and improve food supply worldwide (Wallace 2000). Land and water resources in the Aral Sea basin have been degraded due to the expansion of irrigation networks, on-farm mismanagement of irrigation water and degraded infrastructure. This has caused serious economic and health problems along the lower reaches of the two large rivers flowing into the Aral Sea (Micklin 2000, Dukhovny 2003). Thus, sustainable irrigated agriculture is necessary to maintain food production

*per capita*, but this will require technological improvements in water-use efficiency (WUE).

Regulated deficit irrigation (RDI) consists of finding the optimum balance between water use and crop yield. Under RDI, crop producers allow the crop to experience some water stress, but the water saved should allow an increase in the area irrigated, or it could be put to more productive use such as in industrial activities or for civil uses (English and Raja 1996). ICARDA has shown that a 50 % reduction in irrigation water applied decreased yields by 10–15 %, but overall farm productivity increased by 38 % when the water saved was used on other land (Pereira et al. 2002). However, it is important that farmers have control over the timing of irrigation events and amount of water applied and access to the tools for proper irrigation scheduling. One of these tools is precise



information on factors such as the response and sensitivity of crops to water stress, stages of plant development when deficit irrigation should or should not be performed and how much water can be conserved (Kijne et al. 2003).

Previous studies on RDI were usually performed by either reducing the amount of water that is applied to crops to a fraction of the full evapotranspiration (ET), but otherwise keeping the same frequency of irrigations (Pandey et al. 1984, Oktem et al. 2003, de Souza et al. 2003, Oweis et al. 2004, Chaves et al. 2007), or withholding irrigation at specific growth stages (Nielsen and Nelson 1998, Calvache and Reichardt 1999, Pandey et al. 2000, Boutraa and Sanders 2001, Karam et al. 2007). We chose to impose RDI using increased time intervals between irrigation events, based on the water balance method for irrigation scheduling and greater depletion fractions as proposed by Allen et al. (1998).

Alternate furrow irrigation (AFI) consists of surface irrigation systems supplying water to every second furrow. Several field studies have demonstrated considerably improved WUE with this method (Grimes et al. 1968, Crabtree et al. 1985, Graterol et al. 1993, Kang et al. 2000, Tang et al. 2010). In a controlled environment with a divided root system, Kang et al. (1998) showed that water consumption by maize plants subjected to partial root drying was decreased by 34–37 % while yields only decreased 6–11 %. They also showed that transpiration rate decreased compared to well-watered controls but that the photosynthetic rate and leaf water content remained the same, thus leading to significant increases in WUE, shoot biomass production, and root development and distribution compared to controls. Using a similar method, Kirda et al. (2004) showed no significant decrease in the yield of tomato subjected to half of the irrigation water, when each side received water alternatively. Both groups suggested that by having half of its root system in dry soil, the plant continues to synthesize abscisic acid (ABA) in the roots, which reduces its transpiration rate. However, because water is available, growth is less affected. Dodd et al. (2010) demonstrated that sunflower transpiration was most correlated with the soil matrix potential in the dry side of the pot, closely followed by the ABA concentration in the leaf xylem.

In Uzbekistan, agricultural policies emphasize the culture of cotton, an important component of the Uzbek economy, and to a lesser degree, winter wheat. Both are subject to state regulation through a system of quotas, and little agricultural land is left for other food crops. The value of crop diversification for the economy and the benefit of legumes in cropping systems have long been recognized. Both common bean and mungbean are widely consumed in Central Asia. Common bean (*Phaseolus vulgaris* L.) is the most important pulse crop in the world, being consumed more than any other legume crop. Recent studies suggest that only 7 % of the growing area of common bean worldwide

receives adequate rainfall (Broughton et al. 2003), and 60 % of the production occurs under severe drought stress (Graham and Ranalli 1997). Mungbean, also known as green gram, is a small-seeded legume crop less known in the Americas, but widely cultivated in Asia. It is also cultivated to some degree in the United States and in Australia and is often consumed as sprouts (Lawn and Ahn 1985, Poehlman 1991). The crop is known to perform well under conditions of low soil moisture availability. It remains, however, one of the least researched and most under-exploited legume crops (De Costa et al. 1999).

The main objective of this study was to evaluate the production of food legumes under RDI and AFI. This study is the first to evaluate the effect of RDI using the water balance method and increased time intervals between irrigation events in conjunction with AFI on the yield of common bean and mungbean. Another objective was to compare the performance of common bean and mungbean under Uzbekistan conditions and to determine possible constitutive and adaptive traits. A third objective was to confirm that being short-season crops, legumes could be grown after the harvest of winter wheat (early July) and thus be introduced in the current Uzbek cropping system to improve food security and diversity.

## Materials and methods

### Environment

The experiment was conducted in the Fergana valley (Fig. 1), in Uzbekistan, Central Asia (40°23'N, 71°45'E), from the beginning of July until the onset of cold temperatures in mid-October, in the growing seasons of 2003 and 2004. During this period, the climate is hot and dry, with



**Fig. 1** Map of the Aral Sea basin and location of experimental area (Fergana Valley). The experimental field was situated in the Fergana Valley, Uzbekistan (40°23'N, 71°45'E).

typical daily high temperatures being 35–40 °C, and typical daily lows being 20–25 °C. Rain is infrequent, except in early October. From 15 July to 30 September 2003 and 2004, we recorded a total of 8.8 and 7.6 mm of rainfall, respectively, at our field sites. Climatic data (Fig. 2) were collected using an on-site Vantage Pro Meteorological station (Davis Instruments Corp., Hayward, CA, USA), located approximately 200 m from the field site. Based on textural analyses, soil at the experimental sites was a silt loam in 2003 and ranged from a sandy loam to a silt loam in 2004. The soils had an available water content of 96 mm in 2003 and 75 mm in 2004, in the top 60 cm. The organic matter content was determined in each plot, in the 0- to 30-cm and 30- to 70-cm layers. In both years, the soils had low organic matter contents (less than 2 %) and a well-developed plough pan at 30–40 cm.

### Experimental design

The treatments were organized on the field site following a randomized complete block split-plot design. The treatments were comprised of factorial combinations of three factors: RDI level (recommended level, moderate deficit and severe deficit – see section on Irrigation scheduling for more details), irrigation water distribution pattern (alternate and every furrow irrigation) and crop (common bean and mungbean). RDI treatment was the main plot factor, and the combinations of furrow irrigation strategy and

crop constituted the subplot factor. There were four blocks. Each subplot measured 15 × 12 m with an additional 1.5 m of buffer zone on each side of the irrigation ditch.

### Irrigation scheduling

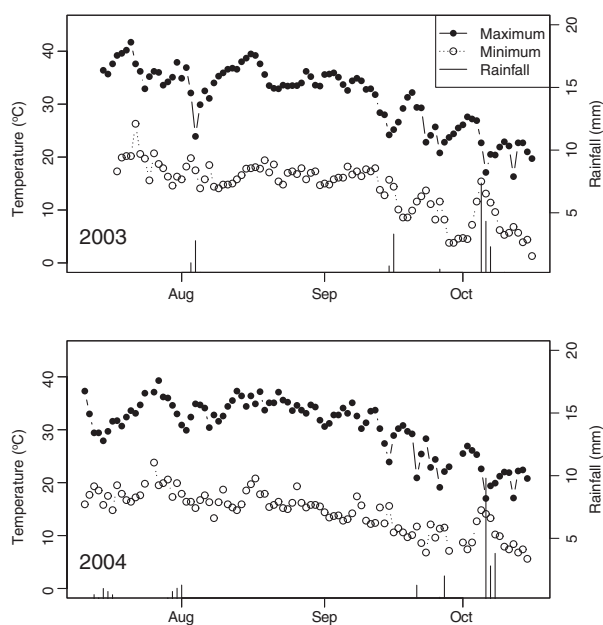
Levels of RDI were determined according to the concept of soil water depletion fractions, as defined by Allen et al. (1998). Depletion fractions are measures of soil water depletion as a percentage of the total available soil water. This results in longer time intervals between irrigation events in treatments with higher depletion fractions. For common bean, the depletion fractions used were 0.45 as the recommended level (Allen et al. 1998) and 0.6 and 0.7 as the moderate and severe stress levels, respectively. For mungbean, the recommended depletion fraction was also 0.45 (Allen et al. 1998), but the moderate and severe stress levels were 0.65 and 0.8, respectively. Local growers informed us that only one irrigation event is often necessary to successfully produce mungbean, suggesting greater tolerance to water stress than most crops. We expected these depletion fractions to be good approximations of a moderate and a severe level of stress for each of the crops. Prior to our work, yield effects of RDI depletion factors for common bean or mungbean were unknown. Soil moisture profiles for all twelve combinations of RDI, AFI and crops are illustrated in Fig. 3. A root depth of 60 cm was assumed based on experience of Dr. M. Horst and later confirmed with root excavations.

Daily ET values were computed from climatic data from our meteorological station (according to the FAO Penman–Monteith equation) and an ET gauge placed in a mungbean plot adjacent to our field (Allen et al. 1998). These were used in six water balances corresponding to six combinations of RDI and crop treatments with soil moisture readings to confirm the actual soil depletion. Once the soil was depleted to the appropriate fraction of soil available water, the irrigation amount was determined from the water balance, and the treatment was irrigated.

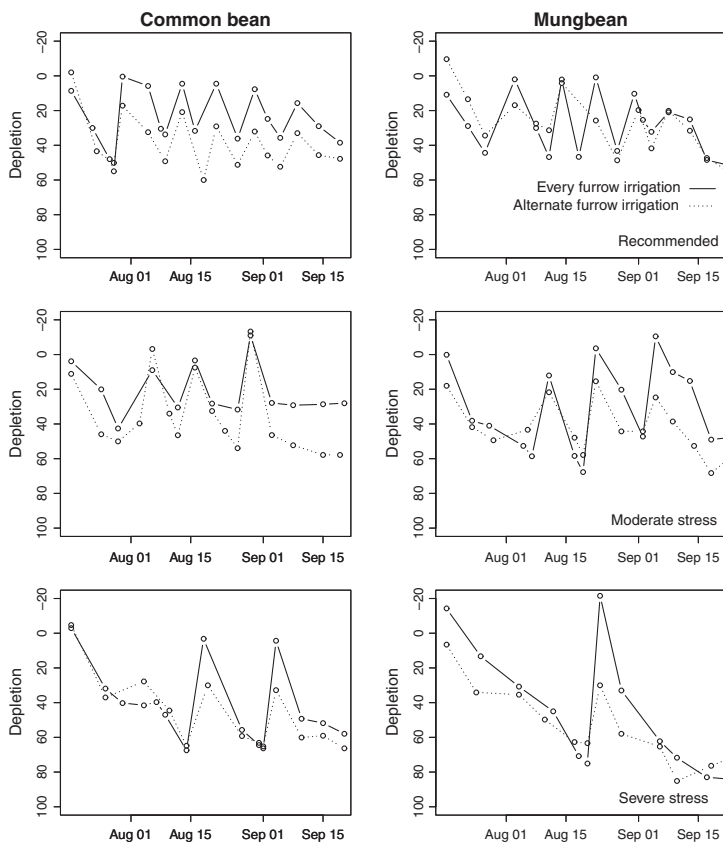
For AFI plots, the head of every second furrow was blocked with mud and straw so that water would not go in this furrow. Blocked furrows were alternated between irrigation events, when there was more than one event. Alternate furrow-irrigated plots were irrigated the same day as the corresponding every furrow-irrigated plot in each crop and RDI level, but only received 75 % of the water applied on a whole-plot basis. More details on the irrigation can be found in Webber et al. (2006).

### Cultural practices

Each field site produced winter wheat immediately prior to our experimentation. The wheat was harvested, the straw



**Fig. 2** Climatic data for the growing seasons of 2003 and 2004 in the Fergana Valley, Uzbekistan (40°23'N, 71°45'E), from the beginning of July until the end of October.



**Fig. 3** Soil moisture profile of the different combinations of regulated deficit irrigation, irrigation strategies and crops under field conditions in the Fergana Valley, Uzbekistan.

and stubble burned and the field ploughed and tractor-drawn levelled, all following standard practices in the region. Sixty-centimetre-wide furrows were formed on the field site with a tractor-drawn lister; a pre-sowing irrigation of approximately  $800 \text{ m}^3 \text{ ha}^{-1}$  was applied to every furrow to consolidate the resulting furrows and raised beds, and to bring the soil back to field capacity.

Seeds of common bean and mungbean were purchased at a local market, and some of the seeds harvested in the first year were kept for planting in the second year. We have retained a sample of these seeds, and they are available, upon request. We have also submitted them to the Australian Tropical Crops and Forages collection and are awaiting accession numbers. Certified cultivars were not available and these landraces were representative of what was available to farmers. The local common bean used in this study had a bushy and determinate growth habit. The local mungbean was also bushy, slightly prone to lodging and semi-determinate. Under the conditions of the field study, both crops were ready to harvest within 90 days. A second flush of flowering did not occur, although the flowering period lasted more than 2 weeks and a few flowers were still observed at the end of the season on mungbean. In addition, as these landraces from Uzbekistan were selected by farmers under hot and dry environments, we

expect this germplasm to exhibit greater drought tolerance than most cultivars of their respective species.

Seeds were sown by hand, at a 5 cm depth, on both sides of the raised beds, at 10-cm intervals, to achieve a plant density of  $300\,000 \text{ plants ha}^{-1}$ . Planting was carried out on 14 and 15 July 2003 and 9–12 July 2004. In 2004, based on experience acquired in 2003, an irrigation event of approximately  $600 \text{ m}^3 \text{ ha}^{-1}$  applied to every furrow was included 5 days after seeding (DAS), to assist in seedling establishment.

Based on the availability of agrochemicals, fertilization in 2003 consisted of superphosphate applied immediately prior to the land levelling, plus a manual application of ammonium phosphate and potassium fertilizer a week after planting at the rate of approximately  $28 \text{ kg ha}^{-1}$  of N,  $51 \text{ kg ha}^{-1}$  of P and  $8 \text{ kg ha}^{-1}$  of K. In 2004, a mix of phosphate and potassium was applied in the furrow immediately prior to the pre-irrigation at a rate of approximately  $40 \text{ kg ha}^{-1}$  for both P and K. Nitrogen fertilizers were not available in 2004. Commercial inoculants for common bean and mungbean were unavailable. Thus, nodule development reported in the results section refers to symbioses formed with indigenous rhizobial strains. Based on the number of nodules and nitrogen analyses of leaves at various stages of growth, it appears that the nitrogen fixation was negligible in these crops. Weed control was conducted manually.

Multiple harvests were necessary as mungbean pods tend to shatter when dry. In 2003, there were two harvests for common bean (3 and 4 October) and six for mungbean (21, 24, 27 and 30 September, and 7 and 13 October). In 2004, there were also two harvests for common bean (20 September and 3 October) and four for mungbean (21 and 26 September, and 1 and 9 October).

### Measurements

An area of 5 × 5 m in the centre of each plot was used for sampling. Mature pods were harvested at regular intervals within the sampling area and threshed by hand. Grain yield was determined for each harvest, then combined across harvests, corrected for moisture content and converted into kg ha<sup>-1</sup> from plant population estimates before statistical analysis.

Stem water potential (SWP) was measured on six plants per plot within the sampling area, 1 day before and 2 days after each irrigation event with a portable pump-up pressure bomb (PMS Instruments, Albany, NY, USA). Stomatal conductance was measured on ten plants, at the same time as SWP, with a diffusion porometer LI-1600M (LICOR Biosciences, Lincoln, NE, USA).

For crop height and number of flowers and pods, six plants were marked at the beginning of the season, and measurements were made on these same plants as the season progressed. In 2003, measurements were taken once a week after crop establishment, while in 2004, measurements were taken twice a week until the first pod harvest. Plants were also harvested at four growth stages during the season for the determination of leaf area, and above-ground biomass dry weight, and root dry weight (the latter in 2004 only). Fifty centimetre of row was harvested, plants were cut off at ground level, and roots were carefully dug out. Leaves were then separated and placed on a white sheet. A digital picture was taken of every individual plant, along with a standard of known area, and these images were later processed with the computer program SigmaScan Pro 5 (Systat Software, San Jose, CA, USA) to determine leaf area. Above-ground biomass was determined from the stem, leaves and reproductive structures when appropriate, of plants harvested in the destructive samplings described earlier. Plants were dried at 70 °C for at least 24 h, to a constant weight. Harvest index was calculated as the ratio of yield over total above-ground biomass at final harvest, all on a dry-weight basis.

### Statistical analyses

Statistical analyses were performed by analyses of variance (ANOVA) and repeated measures analysis by multivariate analyses of variance (MANOVA) using the general linear model (proc GLM) in SAS/STAT software (SAS, Cary, NC,

USA). Following a split-plot design, RDI main effects were tested against the RDI-by-block interaction as the error term. AFI treatment and crops were tested as the subplot effects on the residual error term. Interactions were also tested and presented when significant. We also tested the effect of the 'year' and found significant interactions with the RDI treatments. Therefore, results are presented separately for each year. In general, treatment effects or interactions were considered significant only when they occurred at the 0.05 level of probability. However, in some cases, we considered relevant differences when the probability level was between 0.05 and 0.1; in these cases, the P value is given in the text and in tables when appropriate. If fixed main effects or a fixed interaction was found to be significant in the ANOVA, then mean separations were carried out using *t*-tests on least squares means. In 2003, when significant, the plant population density was used as a covariate. Because there were few nodules, the statistical analysis for nodule number was performed with a nonparametric approach using the RANK procedure prior to proc GLM.

## Results

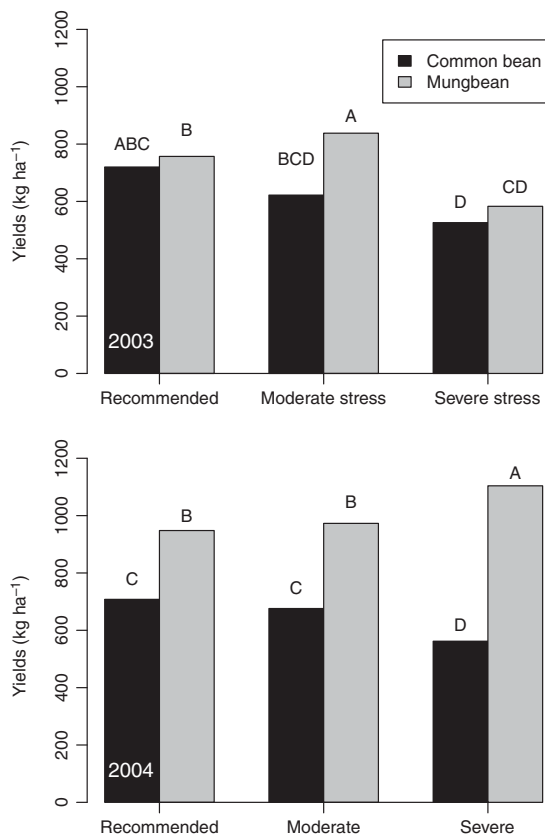
### Yield

The mungbean landrace responded differently than the common bean landrace to the various levels of RDI imposed during this experiment (Fig. 4), resulting in a crop-by-RDI interaction with P values of 0.0913 and 0.0005 for 2003 and 2004, respectively. While common bean yields decreased with increasing stress, mungbean yields were highest at the moderate water stress level (2003) and severe stress level (2004). Common bean yields at the moderate stress level were not significantly lower than bean yields at the recommended irrigation schedule. In addition, AFI did not reduce yields (Table 1), and the interaction with RDI or crop was not significant.

The yield difference between the two experimental years can be explained, at least in part, by the differences in the plant population density. In 2003, variation was relatively high, due to uneven emergence. In particular, plots in the severe stress treatment only reached 50–75 % ground cover, particularly in the mungbean plots. This probably led to yield limitations due to lower radiation interception, but might also have led to greater evaporation from the soil surface, and a lack of transpirational boundary layer. This would have resulted in more stressful environmental conditions for these plants, contributing to lower yields.

### Harvest index

Harvest index (HI) was affected by RDI but differently for the two crops. In the severe stress treatment, HI decreased



**Fig. 4** Seed yields of common bean and mungbean in 2003 and 2004 subjected to regulated deficit irrigation under field conditions in the Fergana Valley, Uzbekistan. Histogram bars with the same letter are not different ( $P < 0.05$ ) within the same year by *t*-tests on least squares means. Each bar represents an average of eight plots from four blocks and two AFI levels. The crop-by-RDI interaction was significant at  $P = 0.0913$  in 2003 and  $P = 0.0005$  in 2004.

in common bean, but increased in mungbean (Fig. 5). In addition, there were no differences between the recommended and moderate stress irrigation schedules within crops. Thus, it seems that mungbean has a greater capacity than common bean to allocate resources to seeds under conditions of severe stress. Again, there were no AFI effects, AFI-by-crop and AFI-by-RDI interactions.

#### Stem water potential

In both years, before irrigation events, mungbean maintained a higher (i.e. less negative) SWP than common bean across RDI and AFI treatments. In addition, SWP was not different between furrow irrigation strategies in either year (Table 1). The response to RDI was different for each year (Table 2). In 2003, there was a strong crop-by-RDI interaction in which mungbean decreased its SWP with increasing stress, while common bean showed the lowest SWP at the

moderate stress level. This interaction, however, was not significant in 2004 ( $P = 0.1238$ ), and numerically, the lowest SWP in common bean occurred in the severe stress level, whereas for mungbean, all depletion levels showed the same SWP.

After irrigation events, the SWP was lower under AFI for both crops and across RDI levels in both years (Table 1), but there was no RDI-by-AFI interaction. Again, the crops responded differently to RDI in each year (Table 2). In 2003, mungbean had a lower SWP after irrigation events in the severe water stress. On the other hand, for common bean, SWP was not different among RDI levels and thus had returned to a relatively high SWP following irrigation. In 2004, mungbean showed the highest SWP after irrigation events and no difference among RDI levels, while common bean had a lower SWP in the severe water-stress treatment. In any case, it seems that both crops are generally able to return to a high SWP after irrigation, no matter how dry the soil was when irrigated.

#### Stomatal conductance

In both years, common bean maintained a higher stomatal conductance than mungbean across RDI and AFI levels before irrigation events. Stomatal conductance of both crops decreased as water stress increased, but decreased proportionally more in common bean than in mungbean, as shown in the crop-by-RDI interaction detected in both years (Table 2).

The stomatal conductance after irrigation events was also higher in common bean than in mungbean across RDI and AFI levels (Table 2). Stomatal conductance was then unaffected by RDI, and values for stomatal conductance were higher than before irrigation events, indicating that gas exchange in all plants was able to recover from the stress, although in 2003, plants in the AFI plots showed a lower stomatal conductance than those in the conventional every furrow irrigation plot (Table 1). This effect, however, was not significant in 2004 ( $P = 0.1410$ ).

#### Crop development

While common bean showed high vigour at the beginning of the season, mungbean grew slowly early in the season, followed by a rapid increase in growth a few weeks after planting, and reached maximum height and biomass near the end of the season. This was clear in the data for above-ground biomass, leaf area and crop height (Fig. 6). Common bean also flowered earlier than mungbean (data partly shown in pod development in Fig. 6d), and as such, the statistical analysis was performed separately for the two crops.



**Table 1** Alternate furrow irrigation as compared to every furrow irrigation on yield, yield components and water relations of common bean and mungbean

Traits	2003			2004		
	Alternate furrow	Every furrow	P value	Alternate furrow	Every Furrow	P value
Yield (kg ha <sup>-1</sup> )	665	686	0.6001	832	826	0.8839
Number of seeds per pod	6.9	7.1	0.1490	6.6	6.8	0.1674
100 seed weight (g)	24.2	24.3	0.9023	19.1	20.0	0.1744
Pods per plant	25.2	26.8	0.1546	9.7	10.2	0.4014
Harvest index (%)	0.302	0.346	0.1339	0.306	0.284	0.2486
Stem water potential (MPa)						
Before irrigation events	-1.02	-1.00	0.5025	-0.81	-0.80	0.0554
After irrigation events	-0.89	-0.82	0.0027	-0.77	-0.75	0.0346
Stomatal conductance (mmol m <sup>-2</sup> s <sup>-1</sup> )						
Before irrigation events	221.0	233.5	0.1409	357.5	330.3	0.0947
After irrigation events	316.9	364.4	<0.0001	449.0	461.7	0.1410
Nitrogen content (%)						
Harvest sampling	2.14	2.21	0.6006	1.77	1.81	0.7824
Grains	3.90	4.03	0.1532	3.93	4.12	0.1412

Values are least-square estimates of means from four blocks, two crops and three levels of RDI treatments, that is, 24 plots when there were no missing observations. There was no AFI level-by-crop or AFI level-by-RDI level interaction in any of the parameters presented.

There was no effect of RDI treatment on above-ground biomass of common bean at any of the growth stages measured in either year. The above-ground biomass of mungbean, however, was lower in the harvest stage for the severe stress treatment in both years. There was also a 45 % decrease in mungbean biomass in 2003 at the flowering stage in the severe stress treatment (data not shown) compared to the recommended schedule, again probably due in part to a plant density-by-RDI interaction in that year, as discussed earlier. Similarly, there were no differences in common bean leaf area due to RDI levels, in either year; however, mungbean leaf area was lower under the severe stress treatment at flowering and harvest in 2003, but in none of the samplings in 2004, although the P value at harvest in 2004 was close to the significance level at 0.0638 (Fig. 6b).

Similarly, there was no effect of RDI on crop height for common bean, but after flowering, mungbean plants in the severe stress treatment were shorter than the other two RDI treatments in both years (Fig. 6c). In general, RDI had no effect on the number of flowers, except for one date in each of the two crops in 2004 only (data not shown). For common bean, 42 DAS, the number of flowers was highest in the moderate stress treatment, but lowest in the severe stress treatment. This might be explained by the fact that the moderate stress treatment had been irrigated a few days before, potentially leading to a flush of flowers following the relief of stress. In mungbean, this difference in the number of flowers occurred 60 DAS, towards the end of flowering. The recommended and moderate stress treatments had more flowers than the severe stress treatment,

indicating that water stress may have shortened the flowering period in mungbean. RDI treatments had no effects on the number of pods m<sup>-2</sup> in either crop in 2004 (Fig. 6d). However, mungbean was affected in 2003, as the number of pods was reduced in the severe stress treatment.

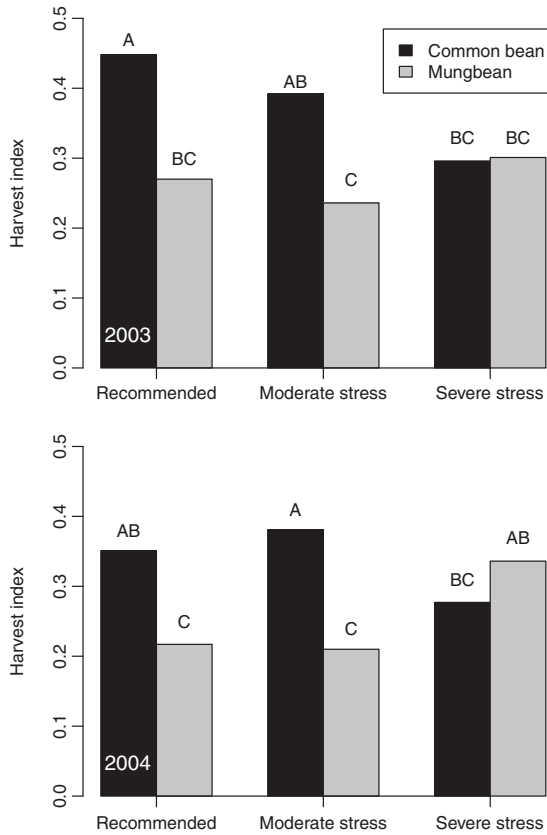
Finally, AFI had no effect on any of the crop development parameters observed (data not shown), and there were no AFI-by-crop or AFI-by-RDI interactions.

### Nodule development

Common bean failed to nodulate with the indigenous rhizobia in the experimental site soil. However, we found nodules in 87 % of mungbean plots in 2003 and 100 % of mungbean plots in 2004, when the plant density was higher. We found an average of 4.2 nodules per plant in 2003 and 6.4 nodules per plant in 2004. However, no differences were observed among RDI levels in either year for the number or dry weight (available only in 2004) of nodules (data not shown).

### Discussion

Our results show that it is not only possible to grow legumes after the harvest of winter wheat, but also that it could be done with one (mungbean) or two (common bean) irrigation events and using AFI. While both crops tolerated some level of stress, mungbean produced the highest yields with the lowest quantity of applied water in 2004. This suggests mungbean is better adapted to the dry and hot conditions of Uzbekistan. Our yield results for



**Fig. 5** Harvest index of common bean and mungbean in 2003 and 2004 subjected to regulated deficit irrigation under field conditions in the Fergana Valley, Uzbekistan. Histogram bars with the same letter are not different ( $P < 0.05$ ) within the same year by *t*-tests on least squares means. Each bar represents an average of eight plots from four blocks and two AFI levels. The crop-by-RDI interaction was significant at  $P = 0.0243$  in 2003 and  $P < 0.0001$  in 2004.

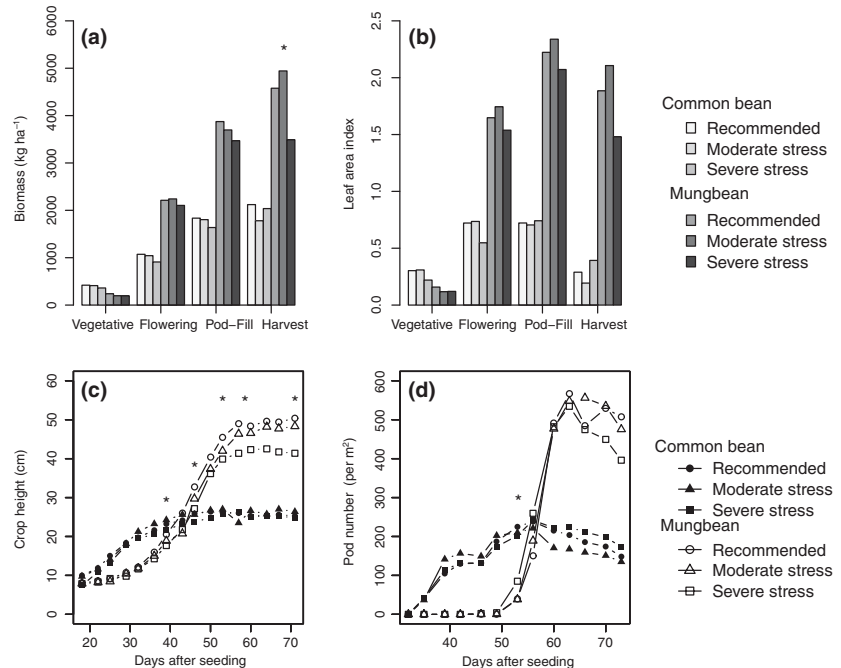
mungbean contradict results from some of the previous experiments comparing the response to irrigation of mungbean with other crops (Pandey et al. 1984, Senthong and Pandey 1989, De Costa et al. 1999, Thomas et al. 2004), but are consistent with the experience of local growers. In a pot experiment, we also found that mungbean yields were reduced by water stress, but to a lesser extent compared to the reduction in yield of common bean (Bourgault et al. 2010). Irrigation schedules comprising four to six irrigation events following seedling establishment have been proposed (Poehlman 1991). In addition, mungbean was found to be quite sensitive to water stress, when compared to a number of other crops, showing the greatest decrease in yield between the well-watered control and most severe stress treatments (Pandey et al. 1984, Senthong and Pandey 1989). These researchers used sprinkler irrigation, and irrigation water was decreased according to the distance of the experimental plot from the sprinkler. The resulting small irrigation depth (water penetration in the soil) may be ineffective for mungbean, which seems to extend its roots deeper into the soil profile to extract water resources from greater depths (Haqqani and Pandey 1994). By filling the root zone during our single irrigation event, a few days before the onset of flowering, we appear to have provided mungbean with sufficient water at a critical time. Angus et al. (1983) found that mungbean was not responsive to irrigation, and Muchow (1985) found that mungbean had the highest yields under water-deficit conditions. In both cases, the dry treatment consisted in no irrigation at all after seedling establishment (and very limited rainfall if any).

Our results regarding the response of common bean to water stress are consistent with the findings of other

**Table 2** Water relations of common bean and mungbean under three levels of regulated deficit irrigation

	Stem water potential (MPa)				Stomatal conductance ( $\text{mmol m}^{-2} \text{s}^{-1}$ )			
	Average before irrigation events		Average after irrigation events		Average before irrigation events		Average after irrigation events	
Crop	2003	2004	2003	2004	2003	2004	2003	2004
Depletion fraction								
Bean								
Recommended	-1.00 bc	-0.91 b	-0.92 c	-0.86 b	317 a	449 a	394 a	569 a
Moderate	-1.17 d	-0.88 b	-0.88 c	-0.90 b	273 b	463 a	388 a	544 a
Severe stress	-1.06 bc	-0.86 b	-0.92 c	-0.87 b	214 c	317 b	389 a	520 a
Mungbean								
Recommended	-0.82 a	-0.70 a	-0.78 b	-0.65 a	209 c	310 bc	282 b	406 b
Moderate	-0.88 b	-0.67 a	-1.06 a	-0.65 a	187 cd	261 d	272 b	354 c
Severe stress	-1.12 cd	-0.70 a	-0.73 c	-0.65 a	164 d	264 cd	300 b	338 c
Interaction significance ( <i>P</i> value)	<0.0001	0.1238	0.0283	0.0223	0.0281	0.0021	0.7492	0.4168

Values are least-square estimates of means from four blocks and two AFI levels. Values with the same letter in the same parameter (i.e. in the same column) are not statistically different at  $P < 0.05$ .



**Fig. 6** Crop development (biomass (a), leaf area index (b), crop height (c) and number of pods (d)) of common bean and mungbean subjected to regulated deficit irrigation under field conditions in the Fergana Valley, Uzbekistan, in 2004. Values are averages of eight plots from four blocks and two AFI levels. Significant differences between RDI treatments for mungbean are identified by an asterisk (\*). No significant differences were observed in common bean.

researchers, who also measured a decrease in yield (Nielsen and Nelson 1998, Dapaah et al. 2000, Boutraa and Sanders 2001, Wakrim et al. 2005). In a companion paper (Webber et al. 2006), we showed that the WUE of common bean remained constant over the various treatments. However, because producers rarely monitor the inflow of irrigation water and typically over-irrigate fields (EC-IFAS 1999), it seems likely that the reduction in the frequency of irrigation events (from four irrigation events to two using a moderate stress RDI schedule) would lead to considerable water savings in a producer's fields. Thus, the potential WUE benefits of RDI in common bean should not be dismissed.

Differences between crops in some of the physiological parameters observed point to what might be important adaptive traits for legumes growing in semi-arid environments. Mungbean was able to maintain or increase its HI under stress, while common bean HI decreased with increasing stress. Its partitioning to yield thus appears to be superior under stress. Mungbean also maintained a lower stomatal conductance across RDI and AFI levels, but a lower proportional decrease with (i.e. a lower response to) increasing stress, which suggests some intrinsic water-saving characteristic rather than an active response to stress. A similar mechanism seems to be at play in cowpea (Hamidou et al. 2007) and Bambara groundnut (Jorgensen et al. 2010), where drought-resistant cultivars showed lower stomatal conductance when well watered. It would be interesting in further studies to investigate whether there is genotypic variability in this trait as well as in epicuticular wax accumulation as found in maize (Meeks

et al. 2012) or in the performance of the antioxidant defence system as found in wheat (Singh et al. 2012) as these mechanisms could be superior in mungbean compared to common bean.

No other physiological or developmental trait can account for the improved performance of mungbean under Uzbekistan conditions. Mungbean also maintained a higher SWP than common bean and showed no response to increasing stress in 2004. Osmotic adjustment (defined as a decrease in osmotic potential to maintain water uptake) in this case does not seem to contribute meaningfully to water-stress tolerance. Similarly, the general lack of differences between irrigation treatments in above-ground biomass weight, leaf area production and the number of flowers and pods further seems to indicate that water stress had very little effects on production *per se*, but rather affected translocation of resources to seeds late in the season.

AFI, which saves 25 % of the water applied by not watering every second furrow, did not reduce yields, or most of the yield components measured, and did not affect crops negatively when combined with RDI treatments. Although SWP was reduced in the AFI treatment after irrigation events, this was not translated into yield differences. Therefore, AFI appears to be a simple yet effective way to increase WUE while maintaining yields. More importantly, this study is the first to evaluate the combination of AFI and RDI, and we have found that AFI does not cause further negative effect on crops already subjected to RDI.

Comparisons of the effects of water-deficit stress on crop yields are difficult due to the various methods used to

impose the water deficit. Previous studies on RDI were usually performed by either reducing the amount of water that is applied to crops to a fraction of the full ET, but otherwise keeping the same frequency of irrigations (Pandey *et al.* 1984, Oktem *et al.* 2003, de Souza *et al.* 2003, Oweis *et al.* 2004, Chaves *et al.* 2007), or withholding irrigation at specific growth stages (Nielsen and Nelson 1998, Calvache and Reichardt 1999, Pandey *et al.* 2000, Boutraa and Sanders 2001, Karam *et al.* 2007). While applying a fraction of the ET might be a practical way to impose RDI with sprinkler and drip irrigation, in surface irrigation systems very small irrigation depths are not technically feasible. Further, these small irrigation depths do not bring the soil profile back to field capacity, but rather wet the upper layers, and result in a soil depletion that increases over time. This could potentially lead to severe damage at the yield formation stage. On the other hand, withholding irrigation water at specific growth stages is too simplistic. Even if rainfall is negligible, air temperature, wind speed, irradiation, relative humidity, crop ground cover, soil's water-holding capacity and fertility conditions are all factors that affect ET and that might vary considerably from year to year. The water stress is thus difficult to reproduce.

The concept of depletion fractions integrates the effects of environmental conditions, crop conditions and management through daily ET and available soil moisture. The depletion fractions are also adjusted for the daily crop ET. RDI using increased time intervals between irrigation events, based on the water balance method for irrigation scheduling and greater depletion fractions (as proposed by Allen *et al.* 1998) and used in Panda *et al.* (2003) and in this study), is probably a better approach because: (i) the method is available no matter what water application technology is being used; (ii) the depletion fractions represent an independent measure of stress over soil types and climatic conditions; (iii) filling the entire soil profile potentially encourages deeper root growth and greater drought tolerance; (iv) the lower frequency of irrigation events lowers the water losses from evaporation; and (v) it is not practical for crop producers to grow a control plot to calculate the ET and then apply a fraction of it to the rest of their fields, as would be required using a fraction of the full ET as used in many research studies.

Although our results show that the combination of RDI and AFI is viable, yields of both mungbean and common bean crops were still relatively low compared to yields achieved in other areas, and the establishment of a genetic improvement programme for these two crops and collaborations with other international centres would be beneficial to the region. It is also important to note that mungbean has not benefited from the same research and breeding efforts as many other legume and cereal crops. Breeding programmes in Thailand and Australia have selected

against undesirable characteristics, such as the shattering of pods, and selected for a more determinate maturing of pods. Some of the developed genotypes might present interesting genetic material for a local breeding programme. Improved germplasm and improved rhizobial inoculants combined with improved distribution could improve agricultural production in Uzbekistan, even within the current cotton and wheat quota policies.

## Conclusions

Overall, RDI in combination with AFI, as well as the cultivation of legumes following the harvest of winter wheat, is not only possible, but could have considerable positive effects on the economy, environment and national food security of Uzbekistan and nearby areas of Central Asia. We suggest that while both mungbean and common bean are possible crops, mungbean is better adapted to hot and dry conditions prevalent in semi-arid areas. Our results suggest that water deficit affects the translocation of resources to seeds and that mungbean is able to maintain its HI under severe stress while common bean did not. Mungbean tolerance to water deficit further seems to be related to lower stomatal conductance across water availabilities. A single, but deep, irrigation event around flowering might be all that is necessary for mungbean to yield well and to do so before the onset of rain and lower temperatures in the fall. RDI is also possible with common bean, where yields were not substantially decreased by the moderate stress treatment. A reduction in irrigation events would also be desirable, as crop producers tend to over-irrigate to ensure even water distribution.

## Acknowledgements

The authors thank the Canadian International Development Agency (CIDA) for funding the field component of this research. We further acknowledge the support of the Fonds Québécois de Recherche sur la Nature et les Technologies (FQRNT) during M. Bourgault's PhD studies, and of the National Science and Engineering Research Council (NSERC) during H. Webber's PhD studies. Thanks are also due to Prof. V. Dukhovny of the Scientific Information Centre of the Interstate Commission for Water Coordination (SIC ICWC) of Central Asia for hosting the Canadian researchers, as well as to all laboratory and field staff of the organization for their help with soil analyses and irrigation scheduling. Thanks are due to R. Baker, C. Senecal and N. Stampfli of the Brace Centre for their support in technical and managerial aspects of the project. We also thank the WASAMED conference for allowing us to present preliminary results from this study.

## References

- Allen, R. G., L. S. Pereira, D. Raes, and M. Smith, 1998: Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements. Food and Agricultural Organization, Rome.
- Angus, J. F., S. Hasegawa, T. C. Hsiao, S. P. Liboon, and H. G. Zandstra, 1983: The water balance of post-monsoonal dryland crops. *J. Agric. Sci.* 1001, 699–710.
- Bourgault, M., C. A. Madramootoo, H. A. Webber, G. Stulina, M. G. Horst, and D. L. Smith, 2010: Effects of deficit irrigation and salinity stress on common bean (*Phaseolus vulgaris* L.) and mungbean (*Vigna radiata* (L.) Wilczek) grown in a controlled environment. *J. Agron. Crop Sci.* 196, 262–272.
- Boutraa, T., and F. E. Sanders, 2001: Influence of water stress on grain yield and vegetative growth of two cultivars of bean (*Phaseolus vulgaris* L.). *J. Agron. Crop Sci.* 187, 251–257.
- Broughton, W. J., G. Hernandez, M. Blair, S. Beebe, P. Gepts, and J. Vanderleyden, 2003: Beans (*Phaseolus* spp.) – model food legumes. *Plant Soil* 252, 55–128.
- Calvache, M., and K. Reichardt, 1999. Effects of water stress imposed at different plant growth stages of common bean (*Phaseolus vulgaris*) on yield and N<sub>2</sub> fixation. In: C. Kirda, P. Moutonnet, C. Hera, and D. R. Nielsen, eds. *Crop Response to Deficit Irrigation*, pp. 121–128. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Chaves, M. M., T. P. Santos, C. R. Souza, M. F. Ortuño, M. L. Rodrigues, C. M. Lopes, J. P. Maroco, and J. S. Pereira, 2007: Deficit irrigation in grapevine improves water-use efficiency while controlling vigour and production quality. *Ann. Appl. Biol.* 150, 237–252.
- Crabtree, R. J., A. A. Yassin, I. Kargougou, and R. W. McNew, 1985: Effects of alternate-furrow irrigation: water conservation on the yields of two soybean cultivars. *Agric. Water Manag.* 10, 253–264.
- Dapaah, H. K., B. A. McKenzie, and G. D. Hill, 2000: Influence of sowing date and irrigation on the yield of pinto beans (*Phaseolus vulgaris*) in a sub-humid tropical environment. *J. Agric. Sci.* 134, 33–43.
- De Costa, W., K. N. Shanmugathasan, and K. Joseph, 1999: Physiology of yield determination of mung bean (*Vigna radiata* (L.) Wilczek) under various irrigation regimes in the dry and intermediate zones of Sri Lanka. *Field Crops Res.* 61, 1–12.
- Dodd, I. C., G. Egea, C. W. Watts, and W. R. Whalley, 2010: Root water potential integrates discrete soil physical properties to influence ABA signalling during partial root drying. *J. Exp. Bot.* 61, 3543–3551.
- Dukhovny, V. A., 2003: The Aral Sea basin – Rumors, realities, prospects. *Irrigation Drainage* 52, 109–120.
- EC-IFAS, 1999: Water Use and Farm Management Survey (WUFMAS) Annual Report 1997. Executive committee of the International Fund for the Aral Sea, TACIS, Brussels.
- English, M., and S. N. Raja, 1996: Perspectives on deficit irrigation. *Agric. Water Manag.* 32, 1–14.
- Graham, P. H., and P. Ranalli, 1997: Common bean (*Phaseolus vulgaris* L.). *Field Crops Res.* 53, 131–146.
- Graterol, Y. E., D. E. Eisenhauer, and R. W. Elmore, 1993: Alternate-furrow irrigation for soybean production. *Agric. Water Manag.* 24, 133–145.
- Grimes, D. W., V. T. Walhood, and W. L. Dickens, 1968: Alternate-furrow irrigation for San Joaquin Valley cotton. *Calif. Agric.* 22, 4–7.
- Hamidou, F., G. Zombre, and S. Braconnier, 2007: Physiological and biochemical responses of cowpea genotypes to water stress under glasshouse and field conditions. *J. Agron. Crop Sci.* 193, 229–237.
- Haqqani, A. M., and R. K. Pandey, 1994: Response of mung bean to water-stress and irrigation at various growth-stages and plant densities: 2. Yield and yield components. *Trop. Agric.* 71, 289–294.
- Jorgensen, S. T., F. Liu, M. Ouédraogo, W. N. Ntundu, J. Sarrazin, and J. L. Christiansen, 2010: Drought responses of two bambara groundnut (*Vigna subterranea* L. Verdc.) landraces collected from dry and humid area of Africa. *J. Agron. Crop Sci.* 196, 412–422.
- Kang, S. Z., Z. S. Liang, W. Hu, and J. H. Zhang, 1998: Water use efficiency of controlled alternate irrigation on root-divided maize plants. *Agric. Water Manag.* 38, 69–76.
- Kang, S. Z., Z. S. Liang, Y. H. Pan, P. Z. Shi, and J. H. Zhang, 2000: Alternate furrow irrigation for maize production in an arid area. *Agric. Water Manag.* 45, 267–274.
- Karam, F., R. Lahoud, R. Masaad, R. Kabalan, J. Breidi, C. Chalita, and Y. Roupheal, 2007: Evapotranspiration, seed yield and water use efficiency of drip irrigated sunflower under full and deficit irrigation conditions. *Agric. Water Manag.* 90, 213–223.
- Kijne, J. W., R. Barker, and D. Molden, 2003. Improving water productivity in agriculture: Editors' Overview. In: J. W. Kijne, R. Barker, and D. Molden, eds. *Improving Water Productivity in Agriculture: Limits and Opportunities for Improvement*, pp. xi–xix. CAB International, Colombo, Sri Lanka.
- Kirda, C., M. Cetin, Y. Dasgan, S. Topcu, H. Kaman, B. Ekici, M. R. Derici, and A. I. Ozguven, 2004: Yield response of greenhouse grown tomato to partial root drying and conventional deficit irrigation. *Agric. Water Manag.* 69, 191–201.
- Lawn, R. J., and C. S. Ahn, 1985: Mung bean (*Vigna radiata* (L.) Wilczek/*Vigna mungo* (L.) Hepper). In: R. J. Summerfield, and E. H. Roberts, eds. *Grain Legume Crops*, pp. 584–623. Collins, London.
- Meeks, M., S. C. Murray, S. Hague, D. Hays, and A. M. H. Ibrahim, 2012: Genetic variation for maize epicuticular wax response to drought stress at flowering. *J. Agron. Crop Sci.* 198, 161–172.
- Micklin, P. P., 2000: *Managing Water in Central Asia*. Royal Institute of International Affairs: Russian and Eurasia programme, London.
- Muchow, R. C., 1985: Phenology, seed yield and water-use of grain legumes grown under different soil-water regimes in a semi-arid tropical environment. *Field Crops Res.* 11, 81–97.



- Nielsen, D. C., and N. O. Nelson, 1998: Black bean sensitivity to water stress at various growth stages. *Crop Sci.* 38, 422–427.
- Oktem, A., M. Simsek, and A. G. Oktem, 2003: Deficit irrigation effects on sweet corn (*Zea mays sacharrata* Sturt) with drip irrigation system in a semi-arid region: I. Water-yield relationship. *Agric. Water Manag.* 61, 63–74.
- Oweis, T., A. Hachum, and M. Pala, 2004: Lentil production under supplemental irrigation in a Mediterranean environment. *Agric. Water Manag.* 68, 251–265.
- Panda, R. K., S. K. Behera, and P. S. Kashyap, 2003: Effective management of irrigation water for wheat under stressed conditions. *Agric. Water Manag.* 63, 37–56.
- Pandey, R. K., W. A. T. Herrera, and J. W. Pendleton, 1984: Drought response of grain legumes under irrigation gradient: 1. Yield and yield components. *Agron. J.* 76, 549–553.
- Pandey, R. K., J. W. Maranville, and A. Admou, 2000: Deficit irrigation and nitrogen effects on maize in a Sahelian environment: I. Grain yield and yield components. *Agric. Water Manag.* 46, 1–13.
- Pereira, L. S., T. Oweis, and A. Zairi, 2002: Irrigation management under water scarcity. *Agric. Water Manag.* 57, 175–206.
- Poehlman, J. M., 1991: *The Mungbean*. Westview Press, Boulder, CO, USA.
- Senthong, C., and R. K. Pandey, 1989: Response of 5 food legume crops to an irrigation gradient imposed during reproductive growth. *Agron. J.* 81, 680–686.
- Singh, S., A. K. Gupta, and N. Kaur, 2012: Differential responses of antioxidative defence system to long-term field drought in wheat (*Triticum aestivum* L.) genotypes differing in drought tolerance. *J. Agron. Crop Sci.* 198, 185–195.
- de Souza, C. R., J. P. Maroco, T. P. dos Santos, M. L. Rodriguez, C. M. Lopez, J. S. Pereira, and M. M. Chaves, 2003: Partial root drying: regulation of stomatal aperture and carbon assimilation in field-grown grapevines (*Vitis vinifera* cv. Moscatel). *Funct. Plant Biol.* 30, 653–662.
- Tang, L. S., Y. Li, and J. H. Zhang, 2010: Biomass allocation and yield formation of cotton under partial rootzone irrigation in arid zone. *Plant Soil* 337, 413–423.
- Thomas, M., M. J. Robertson, S. Fukai, and M. B. Peoples, 2004: The effect of timing and severity of water deficit on growth, development, yield accumulation and nitrogen fixation of mungbean. *Field Crops Res.* 86, 67–80.
- Wakrim, R., S. Wahbi, H. Tahi, B. Aganchich, and R. Serraj, 2005: Comparative effects of partial root drying (PRD) and regulated deficit irrigation (RDI) on water relations and water use efficiency in common bean (*Phaseolus vulgaris* L.). *Agric. Ecosyst. Environ.* 106, 275–287.
- Wallace, J. S., 2000: Increasing agricultural water use efficiency to meet future food production. *Agric. Ecosyst. Environ.* 82, 105–119.
- Webber, H. A., C. A. Madramootoo, M. Bourgault, M. G. Horst, G. Stulina, and D. L. Smith, 2006: Water use efficiency of common bean and green gram grown using alternate furrow and deficit irrigation. *Agric. Water Manag.* 86, 259–268.

## Modeling the influence of tile drainage flow and tile spacing on phosphorus losses from two agricultural fields in southern Québec

J. Morrison, C. A. Madramootoo and M. Chikhaoui

### ABSTRACT

Tile drainage is a widely adopted water management practice in the eastern Canadian provinces of Québec and Ontario. It aims to improve the productivity of poorly drained agricultural fields. Nevertheless, studies have also shown that subsurface drainage is a significant pollution pathway to surface water. This study was undertaken to evaluate the effect of tile drain spacing on surface runoff, subsurface drainage flows, and phosphorus (P) loss from two tile-drained agricultural fields located near Bedford, Québec. Field data were used with the DRAINMOD model, and in developed regression models in order to perform the analysis. Both DRAINMOD and the regression models showed good performance. Simulation results indicated that when lateral tile drain spacing is increased, the volume of subsurface drain flow decreases, and the volume of surface runoff increases, at sites with sandy and clay loam soils. For every 5 m increase in drain spacing, total phosphorus (TP) loads in subsurface drainage decreased by 6% at a site with sandy loam soil, and increased by 20% at a site with clay loam soil. TP loads in surface runoff increased as a result of increased drain spacing.

**Key words** | DRAINMOD, hydrologic modeling, phosphorus management, runoff, tile drainage, water quality

**J. Morrison** (corresponding author)

**C. A. Madramootoo**

Brace Centre for Water Resources Management,  
McGill University – Macdonald Campus,  
3 Stewart Park, 21 111 Lakeshore Road, Ste Anne  
de Bellevue,  
Québec, H9X 3V9,  
Canada  
E-mail: [jane.morrison@mail.mcgill.ca](mailto:jane.morrison@mail.mcgill.ca)

**M. Chikhaoui**

Institute of Agronomy and Veterinary Hassan II,  
PO Box 6608,  
10101 Rabat,  
Morocco

### ABBREVIATIONS

EF	efficiency
ET	evapotranspiration
MAE	mean absolute error
N	nitrogen
P	phosphorus
PET	potential evapotranspiration
RMSE	root mean square error
TDP	total dissolved phosphorus
TP	total phosphorus
TPP	total particulate phosphorus

### INTRODUCTION

Tile drainage improves crop yields and reduces surface runoff and soil erosion, but contributes to the loss of nutrients from agricultural fields (Skaggs & Van Schilfgaarde 1999). As a

result, increased levels of nutrients are transported to downstream freshwater reaches leading to eutrophication (Smith *et al.* 2006). During eutrophication, excess nutrients (such as nitrogen (N) and phosphorus (P)) cause increased aquatic flora biomass production, creating anoxic conditions and increased turbidity. As a result, surface waters experience decreased biodiversity, and poor recreational conditions (Ansari *et al.* 2011). P usually shows limited mobility in soils and its contribution to accelerated water eutrophication is mainly attributed to surface rather than subsurface flow (Sharpley *et al.* 1993). However, in Québec as well as in other Canadian provinces, research investigations have identified subsurface drainage as a significant pollution pathway of P to surface water bodies (Jamieson *et al.* 2003; Eastman *et al.* 2006; Kinely *et al.* 2007).

The partitioning of drainage water between surface and subsurface pathways, and the subsequent potential for surface

runoff have been shown to greatly affect P losses (Enright & Madramootoo 2004). Other studies have clearly shown that the movement of P by subsurface transport depends on the characteristics of the soil profile; P losses through subsurface flow have been shown to depend on the P sorption capacity of soils (Sharpley & Halvorson 1994) as well as the presence or absence of preferential pathways in the soil (Enright & Madramootoo 2004). On the other hand, Enright & Madramootoo (2004) found that there is no relationship between either soil test P, or percent P saturation and P losses in an agricultural field. It is well understood that soil texture largely impacts P losses. Research findings from Chikhaoui *et al.* (2008) found that with a clay loam soil, particulate P is the dominant form of P loss through the subsurface drainage system. In addition, it was found that fine textured soil experiences mostly preferential flow, while coarse textured soil exhibits mostly matrix flow.

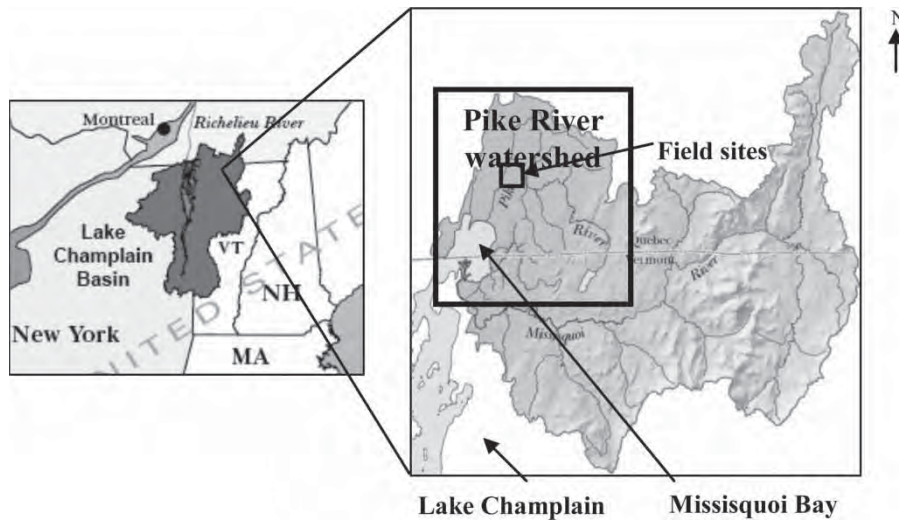
The quantity and quality of water discharged from tile drainage systems are very much dependent on drainage density (Skaggs *et al.* 1995; Chikhaoui *et al.* 2006). Drainage density, defined as the rate of water removed per day, is a function of drain spacing; drainage density is typically high when there is narrow drain spacing. Many studies have quantitatively evaluated the loss of N from subsurface drainage systems using simulation models (Helwig *et al.* 2002; Kladivko *et al.* 2004), indicating that N losses are strongly correlated with drain spacing. Kladivko *et al.* (1999) found that narrow drain spacing had a negative impact on drainage water quality. However, few investigations have evaluated the effect of drain spacing on P losses. In fact, P is now regarded as the limiting nutrient in the development of nutrient management plans in Québec, given the oversaturation of soils by P in Québec.

Simulation models can be a time- and cost-efficient method of developing predictive scenarios by extending existing knowledge. Once validated using field data, a model has the potential to predict outcomes and evaluate scenarios for best management practices under a wide range of conditions (Gollamudi 2006). Several models have been developed to predict subsurface drainage flows and facilitate the design of agricultural drainage systems. DRAINMOD (Skaggs 1978) is widely used in North America to simulate outflows from subsurface tile drainage

systems. This process-based model has been successfully used for different soil types and crop conditions (Helwig *et al.* 2002; Sands *et al.* 2003; Wang *et al.* 2006). However, DRAINMOD does not simulate P loads. To evaluate the effect of drain spacing on P losses, DRAINMOD must be coupled with another model that links subsurface drainage outflows and P loads. Several other models have been developed to simulate P loads, as well as general surface and subsurface agricultural water quality in different regions around the world. Some of these models are: EPIC (Erosion Productivity Impact Calculator; Williams *et al.* 1984), GLEAMS (Groundwater Loading Effects of Agricultural Management Systems model; Leonard *et al.* 1987), WEPP (Water Erosion Prediction Project model; Nearing *et al.* 1989), ADAPT (Agricultural Drainage and Pesticide Transport model; World *et al.* 1988), SWAT (Soil and Water Assessment Tool; Arnold *et al.* 1998) and AnnAGNPS (Annualized Agricultural Nonpoint Source model; Sarangi *et al.* 2007). Every model is unique and requires extensive input data for successful calibration and subsequent use. Some of these models do not account for existing tile drainage systems or snowmelt events that generally occur in cold regions. In order to take these factors into consideration, simple empirical models of P losses through tile drainage can be developed, calibrated with local data, and coupled with hydrological models for use as an alternate predictive tool (Beauchemin *et al.* 2003).

The main objectives of this study were: (1) to calibrate and validate DRAINMOD using observed drainage data from two experimental sites in southern Québec over a period ranging from 2002 to 2004; (2) to establish a relationship between subsurface drain flows and P loads, and to link this empirical model with DRAINMOD outputs; and (3) to investigate the effects of drain spacing on surface runoff, subsurface drainage, and P losses in two selected agricultural tile-drained fields.

This study is based on field measurements at two drain spacings (10 and 13 m), and simulation models were used to extrapolate the results to a range of other drain spacings normally encountered in Québec. The results from this study are limited to the geographic location studied, and similar results may not be attained in an alternate location, or at a different scale.



**Figure 1** | Location of the two experimental sites in the Pike River watershed (adapted from Lake Champlain Basin Program 2007).

## MATERIALS AND METHODS

### Site description

The data used in this study were collected from two experimental agricultural fields located in the Pike River watershed (Québec) about 70 km southeast of Montréal (Figure 1). These two fields are situated on privately owned land and are located approximately 3 km apart. Site A is on a dairy farm with a surface and subsurface drainage area of 6 ha. Site B belongs to a swine and cash crop producer, and has a surface drainage area of 7 ha and a subsurface drainage area of 7.8 ha.

Soil classification of the sites was obtained from local soil surveys. The soil at Site A is a Rubicon sandy loam, exhibiting fair internal drainage with sand and clay contents of 59 and 10%, respectively. The predominant soil type at Site B is a Sainte Rosalie clay loam. The internal drainage of this soil is imperfect; it contains 22% sand and 40% clay. At Site B, deep vertical cracks are common during the summer months when the soil is exposed to periods of extreme wetting and drying. Both soils are characterized by a granular soil structure. Soil test P concentrations and P saturation levels were calculated based on samples collected in 2001 using the Mehlich-III procedure as described by Tran & Simard (1993). A complete list of the soil characteristics is available in Table 1.

These two sites were underlain with a conventional parallel tile drainage system. The subsurface drainage system consists of plastic corrugated lateral pipes of

**Table 1** | Soil phosphorus, tillage, crop rotation, and fertilization in the experimental fields

	Site A		Site B
<b>Soil test P</b>	373 kg ha <sup>-1</sup>		114 kg ha <sup>-1</sup>
<b>P saturation</b>	22%		5.3%
<b>Tillage</b>	Mouldboard plough in the fall (except in 2004)		Mouldboard plough in the fall
<b>Crop rotation</b>			
2001	Corn ( <i>Zea mays</i> L.)		Soybean ( <i>Glycine max</i> L.)
2002	Corn ( <i>Zea mays</i> L.)		Barley ( <i>Hordeum vulgare</i> L.)
2003	Corn ( <i>Zea mays</i> L.)		Corn ( <i>Zea mays</i> L.)
2004	Alfalfa ( <i>Medicago sativa</i> L.)		Corn ( <i>Zea mays</i> L.)
<b>Fertilization</b>	<b>Mth</b>	<b>Phosphorus fertilization</b>	<b>Phosphorus fertilization</b>
2001	May	56 kg ha <sup>-1</sup>	–
2002	May	–	35 kg ha <sup>-1</sup>
	Sept	56 kg ha <sup>-1</sup>	20 kg ha <sup>-1</sup>
2003	May	56 kg ha <sup>-1</sup>	46 kg ha <sup>-1</sup>
2004	May	–	46 kg ha <sup>-1</sup>
	July	20 kg ha <sup>-1</sup>	–
	Oct.	20 kg ha <sup>-1</sup>	–

11 cm diameter and 21 cm diameter outlets. The system was installed with a trenchless plow in a systematic pattern. The outlets were made of corrugated plastic pipe for Site A and clay pipe for Site B. Lateral spacing was 13 m at Site A and 10 m at Site B. The drains were installed at an average depth of 1 m below the soil surface.

Site A was cultivated in corn and alfalfa during the period of study. Site B was cultivated mainly in cash crops such as corn and barley. The cropping sequence for both sites is presented in Table 1. In the study region, the growing season typically lasts from early/mid-May to early/mid-October. Conventional tillage with a mouldboard plough was practiced in the fall at both sites for all study years (2002–2004), except Site A in 2004. The fields were not irrigated. The timing and magnitude of fertilization varied across years of simulation and depended on the crop being cultivated; the quantities of phosphorus fertilizer applied and frequency of application are presented in Table 1. Although the sites were instrumented in October 2000, data from 2001 were unusable as a result of equipment malfunction and were excluded from this study.

### Data collection

The 30-year climatic normal precipitation recorded at Environment Canada's Philipsburg weather station located approximately 9 km from the sites is  $1,096 \text{ mm yr}^{-1}$ , with an annual snowfall equivalent of  $204 \text{ mm yr}^{-1}$ . Rainfall was measured at each site using a tipping bucket rain gage (Texas Instruments TE525M, 0.1 mm tip) and air temperature was recorded using a thermistor (Campbell Scientific 107). The estimated total annual potential evapotranspiration (PET) for the region is  $600 \text{ mm yr}^{-1}$  (Gollamudi *et al.* 2007). This region has an average annual temperature of  $6.8^\circ\text{C}$  and a frost-free period of 155 days (Jamieson *et al.* 2003).

Since 2001, these two experimental fields have been monitored for surface runoff and tile drainage outflows, as well as for various water quality parameters. The instrumentation, data collection, and sampling methodology used during the study period are identical for both sites.

Surface runoff from the fields was measured using H flumes with two water level sensors. The primary sensor was an SR50 ultrasonic depth sensor (Campbell Scientific Inc., Logan, Utah, USA) and the secondary sensor was a

Keller-173 pressure transducer (Campbell Scientific Inc.). Surface runoff volumes were calculated based on a rating curve, specific to the flume specifications, at each site. Flow-proportional composite water samples were collected automatically by American Sigma 900-series water samplers (Hach, Loveland, Colorado, USA).

The subsurface drainage systems were modified to record drain discharge. Two sensors were installed at both sites: the primary sensor was a ProSonic DMU 93 ultrasonic flow meter (Endress and Hauser Canada Ltd, Ontario, Canada) and the secondary sensor was an IF-200 fixed insertion flow meter (Global Water Instrumentation Inc., Gold River, California, USA). Similar to the surface runoff water sampling strategy, flow-proportional water samples were collected automatically by a WS700 composite sampler (Global Water Instrumentation Inc.) from the tile drainage outlet, which is located 1 m below the soil surface. The sampler was activated for every 1 mm of drainage water depth discharged from the tile outlet.

Periodically, grab samples were collected from both surface runoff and subsurface drainage outlets and compared to the other collected composite samples for quality control purposes. Additionally, a hidden duplicate sample was included in each batch for quality control at the lab. Samples were analysed for total phosphorus (TP), total dissolved phosphorus (TDP), and total particulate phosphorus (TPP), as described by Murphy & Riley (1962). The sites were also equipped with soil temperature thermocouples and barometric pressure data loggers, for year-round monitoring. All data collected in the field were checked rigorously for gaps and errors; a complete dataset was available for the two study years.

### Model description

DRAINMOD (version 5.1) is a deterministic hydrologic field scale model developed by Skaggs (1978) to assess the performance of surface and subsurface drainage systems. The model predicts surface runoff, subsurface drainage, evapotranspiration (ET), infiltration, and water table variations. It is conceptualized using a water budgeting protocol for the soil profile section located midway between adjacent drains, and from the impermeable layer up to the soil surface. The model uses Hooghoudt's equation to calculate



subsurface drainage rates, which is based on the Dupuit–Forchheimer assumptions with a correction factor for convergence near the drains (Van Schilfgaarde 1974). When there is surface ponding due to the water table rising to the surface, the drainage rate is estimated using the equation derived by Kirkham (1957). ET is estimated from the PET value, using the Thornthwaite (1948) method. In addition, infiltration is estimated using the Green–Ampt equation and deep seepage is estimated using Darcy’s law. A detailed description of DRAINMOD can be found in Skaggs (1978).

Input data required for DRAINMOD are grouped into subcategories: weather data, drainage design parameters, soil properties, crop information, and trafficability parameters. Additionally, the DRAINMOD 5.1 package includes soil thermal conductivity parameters for simulating field hydrology processes which take place in cold conditions, such as soil freezing, thawing, and snowmelt (Luo *et al.* 2000). Table 2 provides a summary of the hydrologic components and input data used to operate DRAINMOD for the purpose of this study.

It is necessary to calibrate the model with field measurements and to validate it with an independent dataset. After calibrating and validating DRAINMOD, simulations were performed for both sites, analysing different drain spacings ranging from 5 to 70 m, at 5 m increments. This range was chosen to encompass drain spacing values observed in Québec, as well as to provide extreme values for trend analysis.

### Development of an empirical model for P loss prediction

All statistical analyses were performed using IBM SPSS software, version 11 (2002). A threshold level of significance of 0.05 (95% probability level) was used to establish the significance of the different parameters.

The relationships between tile drain outflow and different forms of P loads were estimated using a simple regression analysis. Drain outflow depth obtained as an output of DRAINMOD was used as an input for the regression model. Two regression models were developed for each site describing the relationships between tile drainage flow and TP and TPP. TDP was estimated to be the difference between TP and TPP. The regression models

**Table 2** | Field parameters and DRAINMOD inputs

Parameter	Site A			Site B		
<b>Drainage system</b>						
Field area (ha)	6			7.8		
Drain spacing (m)	13			10		
Drain depth (m)	1			1		
Effective radius of drains (cm)	0.35			0.35		
Drainage coefficient (cm day <sup>-1</sup> )	1			1		
<b>Soil temperature</b>						
Soil thermal conductivity function (coefficients) (W/m °C)	$a = 0.39$			$a = 0.39$		
	$b = 1.326$			$b = 1.326$		
Diurnal phase lag of air temperature (hr)	8			8		
Base temperature as the lower boundary (°C)	7.2			7.2		
Rain/snow dividing temperature (°C)	0			0		
Snowmelt base temperature (°C)	2			2		
Degree day coefficient (mm day <sup>-1</sup> )	5			5		
Critical ice content (cm <sup>3</sup> cm <sup>-3</sup> )	0.2			0.2		
<b>Soil properties</b>						
Bulk density (g cm <sup>-3</sup> )	1.40			1.06		
Organic matter (mean %)	6.81			3.94		
Number of soil horizons in soil profile	3			3		
Soil horizon	1	2	3	1	2	3
Soil horizon thickness (cm)	30	30	40	30	30	40
$K_{\text{sat}}$ (cm hr <sup>-1</sup> )	3.56	1.3	0.9	2.92	0.24	0.14

were developed using drainage water samples collected during the year 2002. The dataset was subdivided into two sets for model development and subsequent validation; out of the total number of samples (47), 75% were used to develop the model and the other 25% were used for evaluating the model’s performance.

Evaluation of the regression models for their accuracy in predicting P loads was carried out using mean absolute error (MAE), and the root mean square error (RMSE). The MAE

and the RMSE were estimated using equations presented by Mayer & Butler (1993):

$$\text{MAE} = \frac{\sum_{i=1}^n |O_i - P_i|}{n} \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (O_i - P_i)^2}{n}} \quad (2)$$

where  $O_i$  is the observed value,  $P_i$  is the predicted value, and  $n$  is the number of values.

The MAE indicates the average absolute difference between predicted and observed values. The RMSE quantifies the relative degree of difference between the model's predicted and observed values. The RMSE also indicates any bias compared to the random variation that may occur (Willmott 1984). If predicted values are exactly equal to observed values, MAE and RMSE are equal to zero.

### Model calibration and evaluation criteria

DRAINMOD was operated using observed data from 2002 to 2004; 2003 and 2004 data were used to calibrate the model, whereas 2002 data were used to validate the model. Observed ET data were not available for the study area; ET values were therefore predicted using the Thornthwaite equation. Throughout the calibration process predicted ET values were optimized by modifying the monthly ET adjustment factors, which were all initially 1.0.

Model-predicted and observed values were compared on a monthly basis. The model performance was evaluated by

calculating modeling efficiency, EF (James & Burges 1982):

$$\text{EF} = \frac{\sum_{i=1}^n (O_i - O)^2 - \sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (O_i - O)^2} \quad (3)$$

where  $O_i$  is the observed value,  $P_i$  is the predicted value,  $n$  is the number of values, and  $O$  is the mean observed value over the time period (1 to  $n$ ).

EF is an overall indication of goodness of fit which directly relates model predictions to observed data. EF evaluates the error relative to the variance of the observed values. Good model performance is indicated by an EF value close to one (Mayer & Butler 1993).

## RESULTS AND DISCUSSION

### Predictions of tile drain outflow

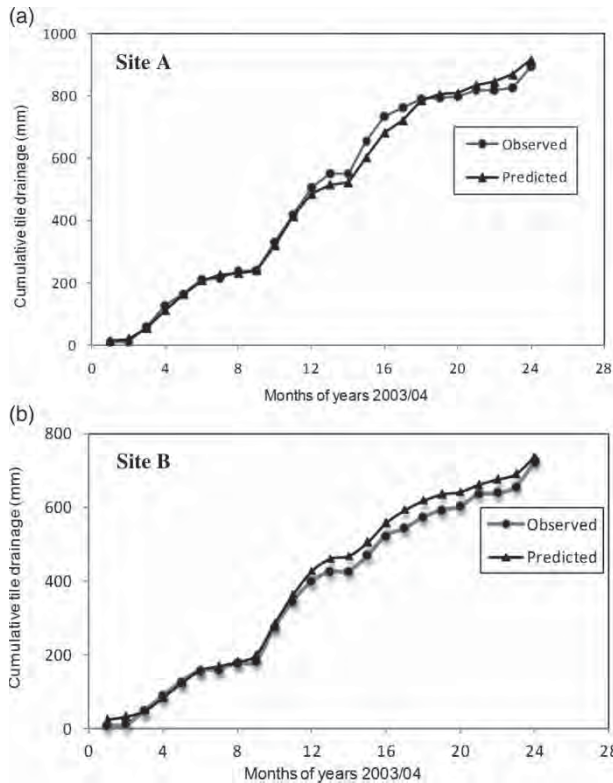
#### Model calibration

Table 3 shows the observed and predicted total yearly subsurface tile drainage volumes based on cumulative monthly data. Figure 2 shows that DRAINMOD accurately predicted tile drainage flows at both sites during the first year of calibration. During the second year of calibration DRAINMOD slightly underestimated drain outflow in the winter and spring, and slightly overestimated drain outflow in the fall at Site A. DRAINMOD slightly overestimated drain outflow during the entire second year of calibration at Site B. The cumulative subsurface drainage outflow depths predicted during the calibration period were 5% higher than the observed values for Site A, and 3% lower

**Table 3** | Comparison of observed and predicted tile drainage flow during calibration and validation periods

Site	Year	Yearly precipitation (mm)	Yearly drainage (mm)		Yearly drainage (% of yearly precipitation)		Average EF
			Observed	Predicted	Observed	Predicted	
A	2003/4 (Calibration)	1,654.6	892.8	946.0	54%	57%	0.69
	2002 (Validation)	912.1	507.0	540.0	56%	59%	0.83
B	2003/4 (Calibration)	1,463.4	721.4	697.2	49%	48%	0.70
	2002 (Validation)	854.8	435.4	401.6	51%	47%	0.70

Note: results are based on cumulative monthly data; EF (efficiency) was calculated using monthly subsurface tile drainage data.



**Figure 2** | Observed and predicted cumulative tile drainage during the calibration period (2003/04) for Site A and Site B.

for Site B. DRAINMOD was very sensitive to adjustments in bulk density and vertical saturated hydraulic conductivity. The variation between the results from Site A and Site B can be attributed to the variability in soil properties, such as texture, bulk density, and saturated hydraulic conductivity at these sites (Table 2). In addition, the overestimation for Site A may be partly due to higher rainfall at this site (Table 3).

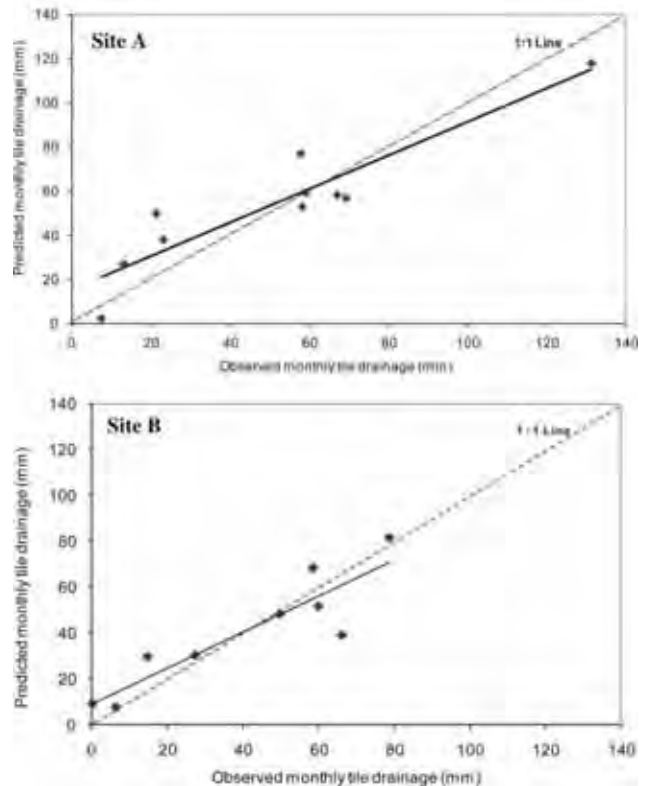
Overall, the slight underestimation or overestimation of monthly subsurface flows can be explained by two factors: first, potential for macropore flow was excluded from the model (Chikhaoui *et al.* 2008) and second, ET was based on adjustment factors in the absence of observed data. Improved ET estimation could improve the performance of the model. It is clear that several other methods would give more accurate ET estimates, such as the Penman equation or the Combination Method, but these would require input data that are difficult to acquire in many scenarios, especially for the long recording periods which are needed to run

DRAINMOD. Nevertheless, the Thornthwaite method is deemed to be sufficiently accurate and appropriate for drainage modeling in humid regions (Skaggs 1980). For these reasons, the Thornthwaite equation was used for ET estimation.

Comparison of these observed and predicted values during the calibration period (2003–2004) at both sites revealed good results. At both sites, yearly average EF values were calculated to be 0.69 or more (Table 3), indicating a satisfactory calibration of DRAINMOD.

### Model validation

Figure 3 illustrates the relationship between monthly observed and predicted values of drain outflow for the validation datasets. At Site A, when approximately 110 mm of tile drainage was predicted, 130 mm was observed. Based on the linear trend, this point appears as though it could be an outlier. However, similar results have been obtained



**Figure 3** | Comparison between observed and predicted monthly tile drainage during the validation period (2002) for Site A and Site B.

**Table 4** | Regression analyses for TP and TPP loads in tile drainage and surface runoff

		Intercept (log(g/ha))			Slope (log(g/ha)/log(mm))			$R^{2a}$	Calculated P
		95% confidence level	Upper limit	Lower limit	95% confidence level	Upper limit	Lower limit		
<b>Site A</b>									
Tile drainage	TP	-0.43	-0.08	-0.78	1.15	1.48	0.82	0.65	51.50
	TPP	-0.26	-0.07	-0.59	0.77	1.08	0.45	0.47	24.60
Surface runoff	TP	0.16	0.42	-0.10	1.74	2.22	1.24	0.65	52.53
	TPP	0.06	0.29	-0.22	1.66	2.18	1.20	0.62	49.93
<b>Site B</b>									
Tile drainage	TP	-0.24	0.03	-0.51	1.52	1.80	1.23	0.78	120.17
	TPP	-0.29	-0.01	-0.57	1.48	1.77	1.18	0.76	105.16
Surface runoff	TP	0.28	0.50	0.05	1.87	2.33	1.40	0.77	69.68
	TPP	0.24	0.05	-0.03	1.73	2.27	1.17	0.67	42.70

<sup>a</sup>Significant at  $\alpha = 0.01$ .

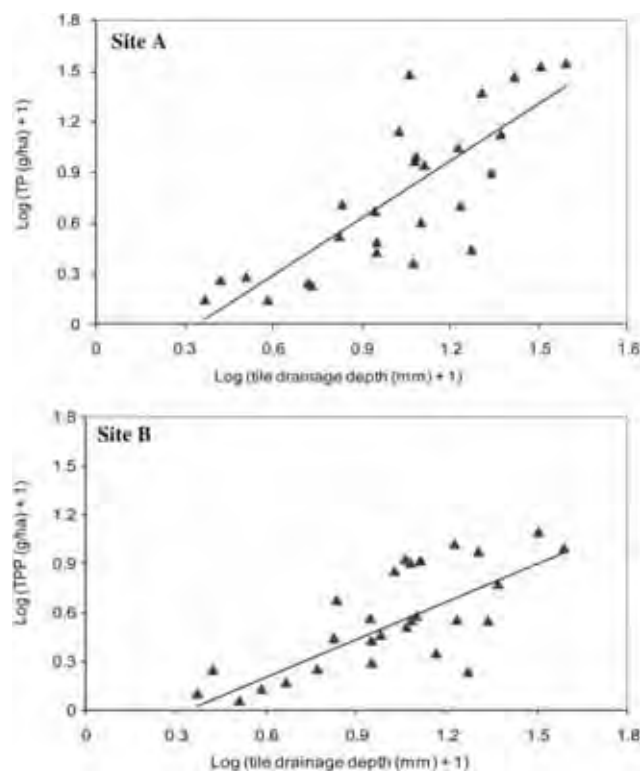
from past research at these two sites (Gollamudi 2006), which confirm that this point should not be considered as an outlier. The regression equation can therefore be used with confidence.

A comparison of observed and predicted values of monthly subsurface tile drainage for the validation period (2002) at both sites revealed satisfactory results with yearly average EF values close to 1 (EF = 0.83 at Site A and EF = 0.70 at Site B) (Table 3). These findings indicate that DRAINMOD performs well in predicting monthly subsurface tile drainage outflows for both soil types for these specific geographical and climatic conditions.

DRAINMOD performed better during validation than during calibration. This can be explained by the fact that validation was carried out during a year in which conditions were ideal for the model performance, i.e., warmer weather with high rainfall. Many large rainfall events occurred during the validation period (2002) and the temperature during that winter was somewhat warmer than during the calibration period (2003 and 2004).

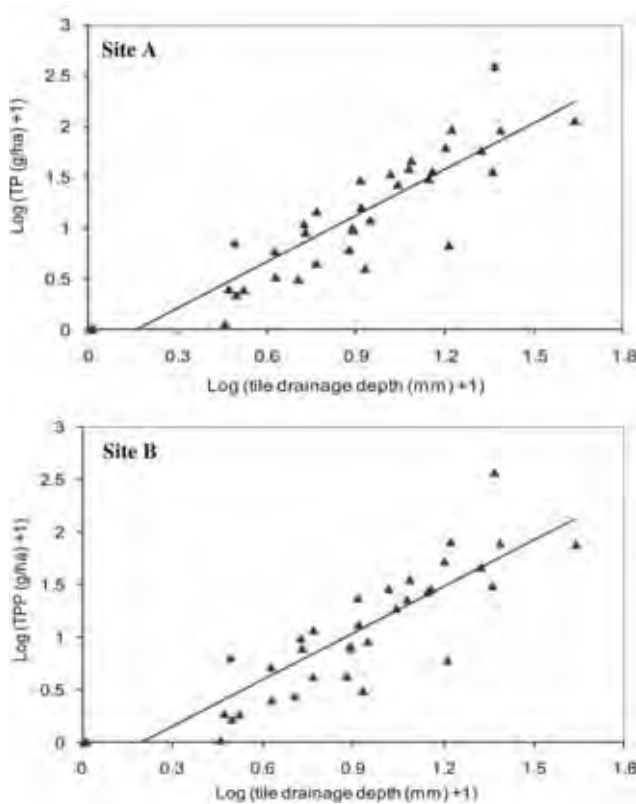
### Regression models for P predictions

The regression models developed for TP and TPP loads in both tile drainage and surface runoff were found to be linear when expressed logarithmically and are expressed in Table 4. The linear relationship between tile drainage depth and TP and TPP loads is graphically displayed for



**Figure 4** | Relationship between the TP and TPP loads in tile drainage and tile drainage depth for Site A.

Sites A and B in Figures 4 and 5, respectively. Table 4 also provides the upper and lower limits of the boundary in which predicted values are found at a 95% confidence level. An analysis of variance test based on simple



**Figure 5** | Relationship between the TP and TPP loads in tile drainage and tile drainage depth for Site B.

regression was performed. It was observed that the calculated P values were larger than the critical P values, meaning that the null hypothesis (slope equal to zero) could be rejected (Table 4). This shows that these values are not significant, and the regression models can be used with 95% confidence.

All P regression models for Site B have greater slopes than for Site A. This indicates increased overall P losses from Site B compared to Site A. Possible reasons for this differentiation are discussed in the subsequent section.

Regression model performance was evaluated by comparing predicted values with observed data. The validation of the regression models was carried out with an independent dataset, i.e., data that were not used to develop the model. For Site A, predicted TP and TPP loads had an MAE of 1.45 or less, and an RMSE of 1.92 or less. For Site B, predicted TP and TPP had MAE values ranging from 2.44 to 6.31 and RMSE values ranging from 2.74 to 9.75 (Table 5). Ideally, MAE and RMSE

**Table 5** | Prediction error statistics of the developed regression models

	Site A				Site B			
	Tile drainage		Surface drainage		Tile drainage		Surface drainage	
	TP	TPP	TP	TPP	TP	TPP	TP	TPP
MAE	1.45	0.85	0.78	0.87	2.44	2.51	6.31	5.92
RMSE	1.92	0.97	0.85	0.96	2.74	3.41	9.75	9.00

values would be equal to zero. In general, MAE and RMSE values were lower for Site A than for Site B. Predictions at Site B were inferior due its problematic soil properties (i.e., soil cracking, preferential flow, high P levels, etc.), whereas Site A exhibited more uniform soil conditions. Nevertheless, predicted values were in sufficiently good agreement with observed values. The developed models are therefore considered appropriate for use in predicting the TP and TPP loads in tile and surface drainage for this experiment.

It is likely that the performance of the regression models would improve if additional variables, like P saturation, soil test P, and other source factors, as well as fertilization rates were included in the model (Gburek *et al.* 2000; Sharpley *et al.* 2001; Beauchemin *et al.* 2003; Chikhaoui *et al.* 2007).

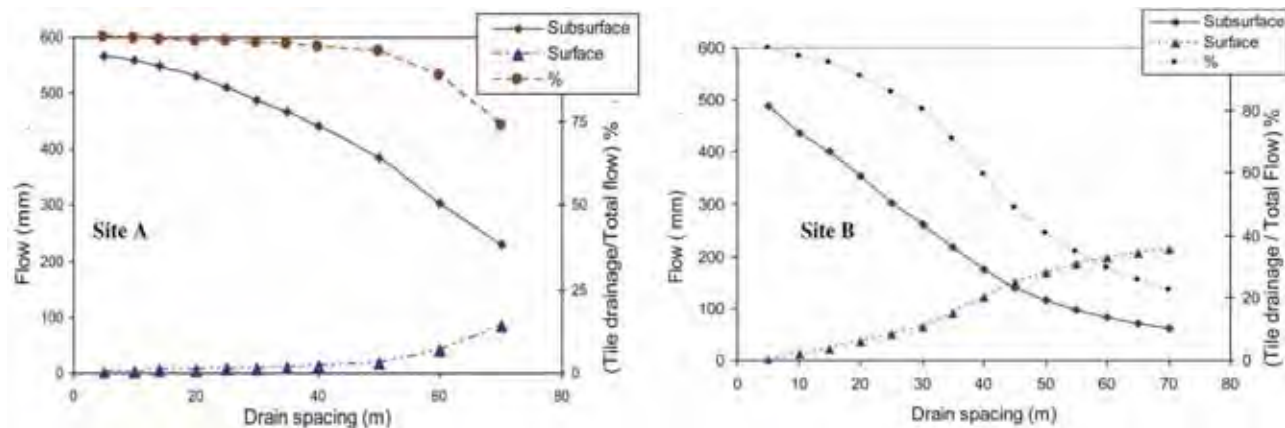
## Effect of drain spacing

### Drainage volumes

The effect of drain spacing on subsurface and surface drainage volumes is shown in Figure 6 for the year 2002. As lateral drain spacing was increased from 5 to 70 m, the depth of subsurface drain flow decreased from 575 to 225 mm at Site A, and from 500 to 60 mm at Site B. Similarly, the depth of surface runoff increased from 0 to 75 mm at Site A, and from 0 to 200 mm at Site B. Thus, subsurface drain outflow decreased and surface runoff increased as drain spacing was increased.

When drain spacing is narrow there is more storage for infiltrating water, thus explaining the reduced surface runoff (Skaggs *et al.* 1982). In most cases, the relationship between drain spacing, surface runoff, and subsurface



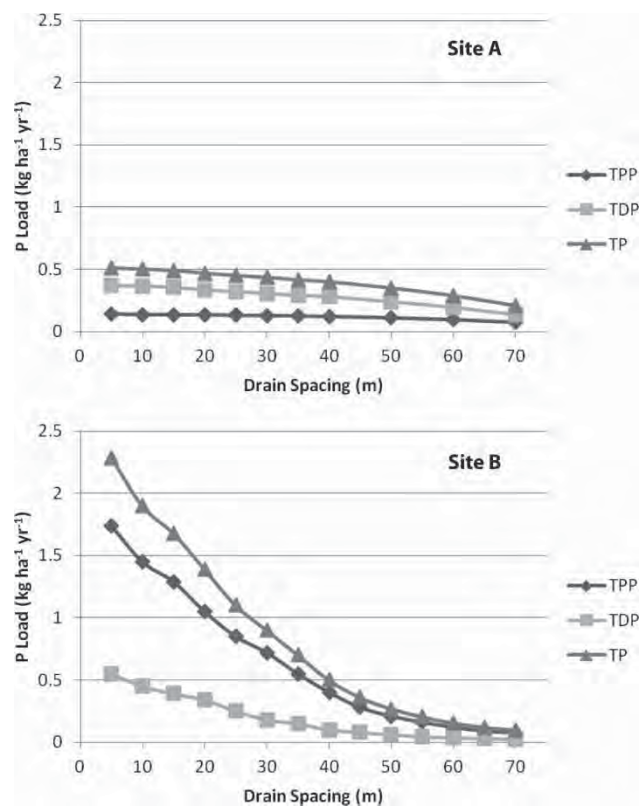


**Figure 6** | Effects of drain spacing on annual tile drain outflows and surface runoff for Site A and Site B.

drainage flows depends on the soil type, or more specifically, soil hydraulic properties such as hydraulic conductivity (Wang *et al.* 2006). The effects of drain spacing on drainage properties were more pronounced for Site B than for Site A. The overall higher subsurface drain flow and lower surface runoff at Site A can be explained by the fact that its soil has a coarse texture, where large pore spaces allow for higher infiltration and therefore larger water removal by the tile drains. Similarly, the greater volume of surface runoff and lower subsurface flow at Site B can be attributed to its finer textured soil, where smaller pore space and non-connecting pores result in lower infiltration (Eastman 2008).

### P losses in tile drainage

The effect of drain spacing on P losses through tile drainage was simulated with a regression model. The simulated results are shown in Figure 7. An increase in drain spacing resulted in a reduction in P loads in drain outflow. For Site A, when drain spacing was increased from 5 to 70 m, TP loads decreased from 0.5 to 0.2 kg ha<sup>-1</sup> yr<sup>-1</sup>. An analysis of the overall results at Site A showed that an incremental increase in drain spacing of 5 m resulted in 6% decrease in TP loads in subsurface drainage. The impact of increasing drain spacing at Site B was more pronounced; when drain spacing was increased from 5 to 70 m, TP loads decreased from 2.3 to 0.1 kg ha<sup>-1</sup> yr<sup>-1</sup>. In this case, an increase of 5 m in drain spacing could result in a 20% decrease in TP



**Figure 7** | Effects of drain spacing on annual TPP, TDP, and TP loads in tile drainage for Site A and Site B.

loads in subsurface drainage. Overall, in these scenarios drain spacing significantly affected P loads in tile drainage.

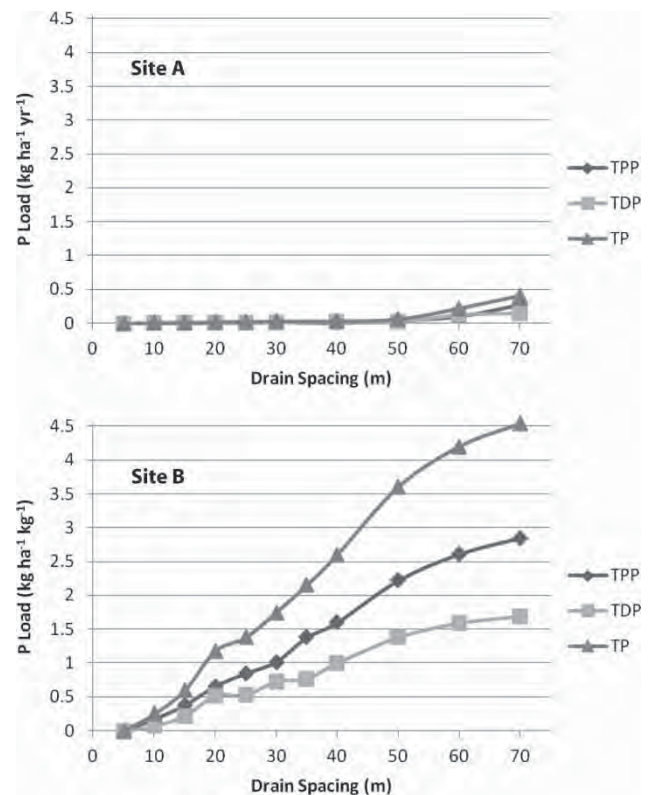
The more significant decrease in TP loads in tile drainage witnessed at Site B is likely due to the fact that at

this site, TPP was dominant in tile drainage, whereas at Site A, TDP was dominant. At Site B, TPP accounted for 84% of TP loads. This predominance may be linked to the mechanisms of P transport which occur under the specific conditions prevalent in the soil profile of a fine textured soil. During dry periods, clay soils tend to shrink, forming deep cracks which serve as conduits for drainage water, thus causing preferential flow (Eastman 2008). This connection between preferential flow and soil cracks was observed at Site B in previous studies (Enright & Madramootoo 2004; Chikhaoui *et al.* 2008). Studies have also concluded that, in clay loam soils, preferential flow is the dominant transport mechanism of soil particulates (Cynor & Findlay 1995); soil macropores can provide a direct route for eroded sediment to subsurface drains. This is corroborated in similar investigations which explore the relationship between sediment load and TPP (Chikhaoui *et al.* 2006). On the other hand, at Site A, TDP accounted for 65% of TP loads. The relatively low TPP loading from Site A can be explained by its coarse soil texture which has the capacity to absorb and store large volumes of water, and which discharges water more gradually (Gollamudi *et al.* 2007). Bypass flow clearly did not occur to the same extent at Site A as it did at Site B. Rather, it is likely that the relatively high rate of TDP in tile drainage at Site A was due to the predominance of matrix flow caused by the soil and water flux interactions, as demonstrated by Chikhaoui *et al.* (2008).

### P losses in surface runoff

The effect of drain spacing on P losses through surface runoff was also simulated with a regression model. The simulated results are shown in Figure 8. An increase in drain spacing resulted in an increase in P loads in surface outflow; when drain spacing was increased from 5 to 70 m, TP loads increased from 0 to 0.4 kg ha<sup>-1</sup> yr<sup>-1</sup> for Site A, and from 0 to 4.5 kg ha<sup>-1</sup> yr<sup>-1</sup> for Site B. As discussed above, when drain spacing is increased, surface runoff also increases. Higher surface runoff results in increased erosion, sediment flow, and thus explains the simulated increase in P losses.

TP loads increased more dramatically for Site B than for Site A. As discussed above, this difference is likely attributed to the finer soil texture at Site B, and dominance of TPP.



**Figure 8** | Effects of drain spacing on annual TPP, TDP, and TP loads in surface runoff for Site A and Site B.

### Optimizing drain spacing

The results from this study clearly demonstrate that drain spacing has a significant effect on P losses through tile drainage and surface runoff. Cumulative TP losses (through both surface runoff and subsurface drainage) from tile-drained fields can be used to evaluate the possibility of reducing total P loadings by optimizing drain spacing at a given site. In order to determine the optimal drain spacing in terms of P loading for a given soil texture and climatic region, other factors such as the impact on crop yield and cost benefits should also be considered. If drains are too close, water is removed from the plant root zone more rapidly, limiting moisture availability and potentially decreasing crop yield. However, drain spacings greater than 40 m would have a negative effect on the agricultural productivity of most soils, since waterlogging of the crops is likely to occur during very high rainfalls (Zhao *et al.* 2000).

## Additional factors influencing drainage and P losses

### Crop rotations

Type of vegetation has an impact on P losses by determining the susceptibility of a field to erosion, as well as the volume of surface runoff. Row crops (such as corn, barley, and soybeans) are not as efficient in terms of reducing surface runoff as hay crops (such as grasses and alfalfa), which are more dense, have a lower soil moisture content due to a higher level of transpiration in the root zone, and have improved infiltration. Furthermore, hay crops have improved soil structure which also facilitate infiltration (Eastman 2008). Site A was planted with corn from 2001 to 2003, while Site B was planted with soybeans, barley, or corn in all years (Table 1). As such, the volume of surface runoff, and P losses at both sites were influenced similarly by the type of crop from 2001 to 2003. In 2004, Site A was under alfalfa while Site B was under corn. Based on this, the infiltration at Site A may have been higher than Site B in 2004, resulting in a lower volume of surface runoff and less erosion at Site A. Therefore, in this year, the influence of the type of crop reinforces the effect that the soil texture at Site A had in decreasing the amount of surface runoff and erosion, and further explaining the overall higher P loads witnessed from Site B.

### Tillage

Tillage practices have an effect on drainage, as well as P losses from agricultural fields. No-till farming conserves soil moisture and improves drainage because of reduced evaporation, increased organic matter, improved soil permeability, and decreased runoff and erosion (Rice 1983). As a result, reduced leaching and runoff of agricultural chemicals (i.e., phosphorus) may occur (Uri 1999). Furthermore, machinery used for tillage could result in soil compaction, leading to poor drainage and increased surface runoff (Eastman 2008). Both sites were consistently tilled with a mouldboard plough in the fall, except for Site A in 2004. The fact that Site A was not tilled in 2004 would imply the possibility for higher subsurface drain flows and decreased P losses in that year. This effect could help to explain the higher drainage flows observed at Site A in 2004. In general, P losses were greater at Site B. The effect

of the tillage regime in 2004, in addition to the soil texture properties at Site B explain this observation.

### Organic matter content

Site A has a higher organic matter content than Site B (Table 2). Organic matter content influences the processes governing nutrient movement in soil and water. When a soil's organic matter content is higher, soil particles are better aggregated into larger units, and there is less soil loss potential (Gollamudi *et al.* 2007). Therefore, Site A would presumably experience less sediment and TPP losses than Site B. This, in combination with all other factors, may help to explain why TDP was dominant in tile drainage at Site A, whereas TPP was dominant in tile drainage at Site B.

### Fertilizer application

Regular fertilizer doses in combination with high rainfall intensity would normally result in high P loading (Eastman *et al.* 2010). On average, Site A received higher fertilization loads than Site B (Table 1). Nevertheless, measured TP losses through tile drainage at Site B were higher than at Site A. Apparently, fertilizer loads at Site A were not so much greater than Site B as to offset the effects of soil texture, and possibly crop type.

### Rain storm and snowmelt events

Rain storm and large snowmelt events increase the possibility for surface runoff. TP concentrations were shown to rise and fall with surface flow rates at both sites. This clearly shows that rain storm and snowmelt events increase TP losses. Similarly, Jamieson *et al.* (2003) found that tile drains are a significant pathway for phosphorus movement during snowmelt events.

Large, intense rainfall events which occurred at the sites throughout the spring and summer (May through August) produced high TP losses. The high percentage of TPP in the subsurface drainage water of Site B during this time period illustrates that even during periods when soil is highly saturated, high levels of TPP losses through the

subsurface drainage system as a result of bypass flow conditions in clay rich soils may still exist (Eastman 2008).

## CONCLUSIONS

Artificial subsurface drainage is commonly used to improve the productivity of poorly drained agricultural soils. However, drainage can have significant impacts on the water quality of downstream rivers and lakes. This study has identified the effects of tile drain spacing on phosphorus losses for two selected tile-drained agricultural fields in Québec. DRAINMOD was linked with two regression models in order to assess P losses in surface runoff and subsurface drainage for different tile drain spacings. DRAINMOD proved to be capable of predicting monthly subsurface drainage outflow satisfactorily. Regression models were developed in order to predict TP, TDP, and TPP loads. This approach, involving a minimal number of parameters, predicted P loads with acceptable accuracy. Simulated results showed that, when climate, geography, and field management practices are kept constant, an incremental increase of 5 m in tile drain spacing could result in a 20% decrease in TP loads in subsurface drainage for a fine textured soil. A lesser impact was observed for a coarse textured soil; where an incremental increase of 5 m in drain spacing resulted in a 6% decrease in TP loads in subsurface drainage. Generally, the simulated results indicated that a significant reduction of P loads in drainage systems occurs when tile drain spacing is increased. Soil texture and structure are extremely important considerations when assessing the risk of P loss from an agricultural field. Determining optimal tile drain spacing is complex because agronomic benefits, environmental impacts, and cost benefits must be balanced. Additional field research is needed to evaluate the effects of drainage system parameters on surface runoff, subsurface drainage, and P losses, with a focus on possible field management interventions.

## ACKNOWLEDGEMENTS

This research was funded by the Fonds Québécois de la Recherche sur la Nature et les Technologies (FQRNT) and

by the Max Bell Foundation. The authors thank the many people who have contributed to this study over the years: Mr Apurva Gollamudi (McGill University), Mr Peter Enright (McGill University), Mr Richard Lauzier (Ministère de l'Agriculture, des Pêcheries et de l'Alimentation du Québec), and Mr Martin Mimeault (Ministère du Développement Durable, de l'Environnement et des Parcs). We also express our sincere appreciation to the Institut de Recherche et de Développement en Agroenvironnement for their technical assistance and involvement with the water sample analysis. We thank the two land owners who allowed us to use their fields for this study.

## REFERENCES

- Ansari, A. A., Gill, S. S. & Khan, F. A. 2011 Eutrophication: threat to aquatic ecosystems. In: *Eutrophication: Causes, Consequences and Control* (A. A. Ansari, S. S. Gill, G. R. Lanza & W. Rast, eds). Springer, New York, pp. 143–170.
- Arnold, J. G., Srinivasan, R., Mutticah, R. S. & Williams, J. R. 1998 [Large area hydrologic modeling and assessment part I: Model development](#). *J. Am. Water Resour. Assoc.* **34** (1), 73–89.
- Beauchemin, S., Simard, R. R., Bolinder, M. A., Nolin, M. C. & Cluis, D. 2003 Prediction of phosphorus concentration in tile-drainage water from the Montréal lowlands soils. *Can. J. Soil Sci.* **83**, 73–87.
- Chikhaoui, M., Madramootoo, C. A. & Stämpfli, N. 2006 Effect of tile drain spacing on phosphorus losses from agricultural fields. Presented at the 22nd Eastern Canadian Symposium of the Canadian Association on Water Quality, Montréal, Québec, Canada, November 3.
- Chikhaoui, M., Madramootoo, C. A. & Stämpfli, N. 2007 Gestion du phosphore et développement d'un outil d'aide à la décision: Indice de phosphore. Presented at the 75th Meeting of ACFAS, Trois-Rivières, Québec, Canada, May 9.
- Chikhaoui, M., Madramootoo, C. A., Eastman, M. & Michaud, A. 2008 Estimating preferential flow to agricultural tile drains. Presented at the ASABE Annual International Meeting, Providence, Rhode Island, USA, June 29–July 2.
- Cynor, J. D. & Findlay, W. I. 1995 Soil and phosphorus loss from conservation and conventional tillage in corn production. *J. Environ. Qual.* **12**, 734–741.
- Eastman, M. 2008 *Field-scale Nutrient Transport Monitoring and Modeling of Subsurface and Naturally Drained Agricultural Lands*. MSc Thesis, McGill University, Montréal, Québec, Canada.
- Eastman, M., Gollamudi, A., Stämpfli, N., Madramootoo, C. A. & Sarangi, A. 2010 [Comparative evaluation of phosphorus losses from subsurface and naturally drained agricultural](#)



- fields in the Pike River watershed of Québec, Canada. *Agr. Water Manage.* **97**, 596–604.
- Eastman, M., Madramootoo, C. A., Stämpfli, N. & Chikhaoui, M. 2006 Surface and subsurface phosphorus losses from agricultural fields in Québec. Presented at the 22nd Eastern Canadian Symposium of the Canadian Association on Water Quality, Montréal, Québec, Canada, November 3.
- Enright, P. & Madramootoo, C. A. 2004 Phosphorus losses in surface runoff and subsurface drainage waters on two agricultural fields in Québec. In: *Proceedings of the Eighth International Drainage Symposium*. American Society of Agricultural Engineers, Sacramento, California, USA, March 21–24.
- Gburek, W. J., Sharpley, A. N., Heathwaite, L. & Folmar, G. J. 2000 Phosphorus management at the watershed scale: a modification of the phosphorus index. *J. Environ. Qual.* **29**, 130–144.
- Gollamudi, A. 2006 *Hydrological and Water Quality Modeling of Agricultural Fields in Québec*. M. Sc. Thesis, McGill University, Montréal, Québec, Canada.
- Gollamudi, A., Madramootoo, C. A. & Enright, P. 2007 Water quality modeling of two agricultural fields in southern Québec using SWAT. *Trans. Am. Soc. Agric. Eng.* **50**, 1973–1980.
- Helwig, T. G., Madramootoo, C. A. & Dodds, G. T. 2002 Modeling nitrate losses in drainage water using DRAINMOD 5.0. *Agr. Water Manage.* **56**, 153–168.
- James, L. D. & Burges, S. J. 1982 Selection, calibration and testing of hydrologic models. In: *Hydrologic Modeling of Small Watersheds* (C. T. Haan, H. P. Johnson & D. L. Brakensiek, eds). ASAE Monograph No. 5, American Society of Agricultural Engineers, St Joseph, MI, USA, pp. 437–472.
- Jamieson, A., Madramootoo, C. A. & Enright, P. 2003 Phosphorus losses in surface and subsurface runoff from a snowmelt event on an agricultural field in Québec. *Can. Biosyst. Eng.* **45**, 11–17.
- Kinely, R. D., Gordon, R. J., Stratton, G. W., Patterson, G. T. & Hoyle, J. 2007 Phosphorus losses through agricultural tile drainage in Nova Scotia, Canada. *J. Environ. Qual.* **36**, 469–477.
- Kirkham, D. 1957 Theory of seepage of ponded water into drainage facilities. In: *Drainage of Agricultural Lands* (J. N. Luthin, ed.). Agronomy Monograph 7, American Society of Agronomy & Soil Science Society of America, Madison, WI, USA, pp. 139–181.
- Kladivko, E. J., Frankenberger, J. R., Jaynes, D. B., Meek, D. W., Jenkinson, B. J. & Fausey, N. R. 2004 Nitrate leaching to subsurface drains as affected by drain spacing and changes in crop production system. *J. Environ. Qual.* **33**, 1803–1813.
- Kladivko, E. J., Grochulska, J., Turco, R. F., Van Scoyoc, G. E. & Eigel, J. D. 1999 Pesticide and nitrate transport into subsurface tile drains of different spacings. *J. Environ. Qual.* **28**, 997–1004.
- Leonard, R. A., Knisel, W. G. & Still, D. A. 1987 GLEAMS: groundwater loading effects on agricultural management systems. *Trans. Am. Soc. Agric. Eng.* **30** (5), 1403–1428.
- Luo, W., Skaggs, R. W. & Chescheir, G. M. 2000 DRAINMOD modifications for cold conditions. *Trans. Am. Soc. Agric. Eng.* **43** (6), 1569–1582.
- Mayer, D. G. & Butler, D. G. 1993 Statistical validation. *Ecol. Modell.* **68**, 21–32.
- Murphy, J. & Riley, J. R. 1962 A modified single solution method for determination of phosphates in surface waters. *Analyt. Chim. Acta* **27**, 31–36.
- Nearing, M. A., Foster, G. R., Lane, L. J. & Finkner, S. C. 1989 A process-based soil erosion model for USDA water erosion prediction project technology. *Trans. Am. Soc. Agric. Eng.* **32** (5), 1587–1593.
- Rice, R. W. 1983 *Fundamentals of No-Till Farming*. American Association for Vocational Instructional Materials, Athens, GA, USA.
- Sands, G. R., Jin, C. X., Mendez, A., Basin, B., Wotzka, P. & Gowda, P. 2003 Comparing the subsurface drainage flow prediction of the DRAINMOD and ADAPT models for a cold climate. *Trans. Am. Soc. Agric. Eng.* **46**, 645–656.
- Sarangi, A., Cox, C. A. & Madramootoo, C. A. 2007 Evaluation of the AnnAGNPS model for prediction of runoff and sediment yields in St. Lucia watersheds. *Biosyst. Eng.* **97**, 241–256.
- Sharpley, A. N. & Halvorson, A. D. 1994 The management of soil phosphorus availability and its impact on surface water quality. In: *Advances in Soil Sciences. Soil Processes and Water Quality* (R. Lal & B.A. Stewart, eds). Lewis Publishers, Boca Raton, FL, USA, pp. 7–90.
- Sharpley, A. N., Daniel, T. C. & Edwards, D. R. 1993 Phosphorus movement in the landscape. *J. Prod. Agric.* **6**, 492–500.
- Sharpley, A. N., McDowell, R. W. & Kleinman, P. J. A. 2001 Phosphorus loss from land and water: Integrating agricultural and environmental management. *Plant Soil* **237**, 287–307.
- Skaggs, R. W. 1978 A water management model for shallow water table soils. Technical report 134 of the Water Resources Research Institute of the University of North Carolina. North Carolina State University, Raleigh, NC, USA.
- Skaggs, R. W. 1980 DRAINMOD reference report. Methods for design and evaluation of drainage-water management systems for soils with high water tables. United States Department of Agriculture, Soil Conservation Service, Fort Worth, TX, USA.
- Skaggs, R. W. & Van Schilfgaarde, J. 1999 *Agronomy Monograph 38, Agricultural Drainage*. American Society of Agronomy, Madison, WI, USA.
- Skaggs, R. W., Brevé, M. A., Mohammad, A. T., Parsons, J. E. & Gilliam, J. W. 1995 Simulation of drainage water quality with DRAINMOD. *Irrig. Drain. Syst.* **9**, 259–277.
- Skaggs, R. W., Nassehzadeh-Tabrizi, A. & Foster, G. R. 1982 Subsurface drainage effects on erosion. *J. Soil Water Conserv.* **37**, 167–172.
- Smith, V. H., Joye, S. B. & Howarth, R. W. 2006 Eutrophication of freshwater and marine ecosystems. *Limnol. Oceanogr.* **51**, 351–355.
- Thornthwaite, C. W. 1948 An approach toward a rational classification of climate. *Geol. Rev.* **38**, 55–94.



- Tran, T. S. & Simard, R. R. 1993 Mehlich III – Extractable elements. In: *Soil Sampling and Methods of Analysis* (M. R. Carter, ed.). Lewis Publishers, Boca Raton, FL, USA, pp. 43–49.
- Uri, N. D. 1999 *Conservation Tillage in U.S. Agriculture: Environmental, Economic, and Policy Issues*. Food Products Press, Binghamton, NY, USA.
- Van Schilfgaarde, J. (ed.) 1974 *Nonsteady flow to drains*. In: *Drainage for Agriculture*. American Society of Agronomy, Madison, WI, USA, pp. 245–307.
- Wang, X., Mosley, C. T., Frankenberger, J. R. & Klavivko, E. J. 2006 [Subsurface drain flow and crop yield predictions for different drain spacings using DRAINMOD](#). *Agric. Water Manage.* **79** (2), 113–136.
- World, A. D., Alexander, C. A., Fausey, N. R. & Dorsey, J. D. 1988 The ADAPT agricultural drainage and pesticide transport model. In: *Modeling Agricultural, Forest and Rangeland Hydrology: Proceedings of the 1988 International Symposium*. American Society of Agricultural Engineers, St Joseph, MI, USA, pp. 129–141.
- Williams, J. R., Jones, C. A. & Dyke, P. T. 1984 A modeling approach to determining the relationship between erosion and soil productivity. *Trans. Am. Soc. Agric. Eng.* **27**, 129–144.
- Willmott, C. J. 1984 On the evaluation of model performance in physical geography. In: *Spatial Statistics and Models* (G. L. Gaile & C. J. Willmott, eds). D. Reidel, Hingham, MA, USA, pp. 443–460.
- Zhao, S. L., Gupta, S. C., Huggins, D. R. & Moncrief, J. F. 2000 [Predicting subsurface drainage, corn yield, and nitrate nitrogen losses with DRAINMOD-N](#). *J. Environ. Qual.* **29**, 817–825.

First received 28 September 2012; accepted in revised form 23 April 2013

# Modeling the Impacts of Spatial Heterogeneity in the Castor Watershed on Runoff, Sediment, and Phosphorus Loss Using SWAT: I. Impacts of Spatial Variability of Soil Properties

Alaba Boluwade · Chandra Madramootoo

Received: 1 April 2013 / Accepted: 8 August 2013 / Published online: 8 September 2013  
© The Author(s) 2013. This article is published with open access at Springerlink.com

**Abstract** Spatial accuracy of hydrologic modeling inputs influences the output from hydrologic models. A pertinent question is to know the optimal level of soil sampling or how many soil samples are needed for model input, in order to improve model predictions. In this study, measured soil properties were clustered into five different configurations as inputs to the Soil and Water Assessment Tool (SWAT) simulation of the Castor River watershed (11-km<sup>2</sup> area) in southern Quebec, Canada. SWAT is a process-based model that predicts the impacts of climate and land use management on water yield, sediment, and nutrient fluxes. SWAT requires geographical information system inputs such as the digital elevation model as well as soil and land use maps. Mean values of soil properties are used in soil polygons (soil series); thus, the spatial variability of these properties is neglected. The primary objective of this study was to quantify the impacts of spatial variability of soil properties on the prediction of runoff, sediment, and total phosphorus using SWAT. The spatial clustering of the measured soil properties was undertaken using the regionalized with dynamically constrained agglomerative clustering and partitioning method. Measured soil data were clustered into 5, 10, 15, 20, and 24 heterogeneous regions. Soil data from the Castor watershed which have been used in previous studies was also set up and termed

“Reference”. Overall, there was no significant difference in runoff simulation across the five configurations including the reference. This may be attributable to SWAT's use of the soil conservation service curve number method in flow simulation. Therefore having high spatial resolution inputs for soil data may not necessarily improve predictions when they are used in hydrologic modeling.

**Keywords** Heterogeneities · Hydrology · Model averaging · Spatial variability · Water quality

## 1 Introduction

Spatial heterogeneity of soil properties in hydrologic and water quality models is often ignored or rarely taking into consideration in the management of freshwater resources (Steinman and Denning 2005). Heterogeneity of soil and land processes contributes to all aspects of the hydrologic cycle (Tague 2005). Having an understanding of the types and processes of spatial heterogeneity is a fundamental component of both theoretical and applied hydrology (Tague 2005). In other words, incorporation of spatial heterogeneity into hydrologic and water quality models should provide a new way to view water quality prediction and freshwater management. This should lead to potentially useful management strategies (Steinman and Denning 2005).

In water quality predictions, there is a need for good nutrient management plans, along with monitoring and decision support systems for nonpoint sources control. Daily or weekly sampling techniques of soil and water

---

A. Boluwade (✉) · C. Madramootoo  
Department of Bioresource Engineering,  
Macdonald Stewart Building, McGill University,  
2111 Lakeshore Road, Sainte-Anne-de-Bellevue, Quebec  
H9X 3V9, Canada  
e-mail: alaba.boluwade@mail.mcgill.ca

variables are technically feasible but not economical. Therefore, there is a need for good prediction models which take spatial watershed variability into account to aid in this circumstance.

One of the most widely applied hydrologic and water quality models is the Soil and Water Assessment Tool (SWAT), developed by the US Department of Agriculture (Arnold et al. 1998). SWAT needs spatial inputs such as a digital elevation model (DEM) as well as land use and soil maps. The modeling procedure and various applications can be found in Arnold et al. (1998) and Gassman et al. (2007). There is a major challenge in the application of this model: the representation of soil heterogeneities. The Hydrologic Response Unit (HRU) is a virtual but unique homogeneous classification of units (areas) of similar land-use, soil class and slope. These units are then aggregated together to form a sub-basin. Each sub-basin drains into a stream which exits to a main drainage waterway. The assumption is that spatial distributions of soil properties are preserved in HRUs; however, in the soil database, the soil classes use mean values for all soil properties. Therefore, the spatial variations of these properties are not completely represented in SWAT.

Given the foregoing, the aim of this research is to quantify the impact of incorporating spatial variability and heterogeneity of soil properties into SWAT, specifically in terms of predicting runoff, sediment, and total phosphorus (TP) loss from agricultural fields. This paper is the first of two regarding the quantification of the impacts of soil heterogeneities on prediction of runoff, sediment, and P loss to Missisquoi Bay. The second paper will involve the quantification of heterogeneities (of other watershed features such as land use) and the sensitivities of SWAT configurations to long-term predictions of runoff, sediment, and TP loss.

## 2 Materials and Methods

### 2.1 Study Watershed Description

Castor watershed (45°08'N, 72°58'W, 11-km<sup>2</sup> area), a sub-watershed of the Pike River, is located near the town of Bedford in southern Quebec, Canada (Fig. 1). The Pike River watershed drains into Missisquoi Bay (north-eastern portion of Lake Champlain). There is a great deal of evidence that the eutrophication seen in the Bay is due to excess nutrient losses from agricultural watersheds

(Hegman et al. 1999; Jamieson et al. 2003; Deslandes et al. 2004; Medalie and Smeltzer 2004). This is a serious challenge from a water quality perspective.

Castor watershed's landscape ranges from 36 to 49 m of elevation above sea level. There is a transition gradient from low, poorly drained lacustrine clays to loamy glacial calcareous tills (Michaud 2004). This result in low water infiltration and poor drainage, which along with wet antecedent moisture conditions and snow melt have contributed to flooding events in recent years. Consequently, subsurface drains are installed on most farms to remove excess water from agricultural fields. The water table in this region is reported to be shallow (Deslandes et al. 2007). The mean long-term annual precipitation (1971–2007) of this study area is about 1,057 mm (Michaud 2004). The watershed is an area of intensive agricultural activity, where land is cropped 44 % to corn (*Zea mays* L.), 28 % to grass (hay), and 20 % to cereals. Other agricultural operations such as swine, poultry, and dairy production make up the remaining 8 % (Deslandes et al. 2007).

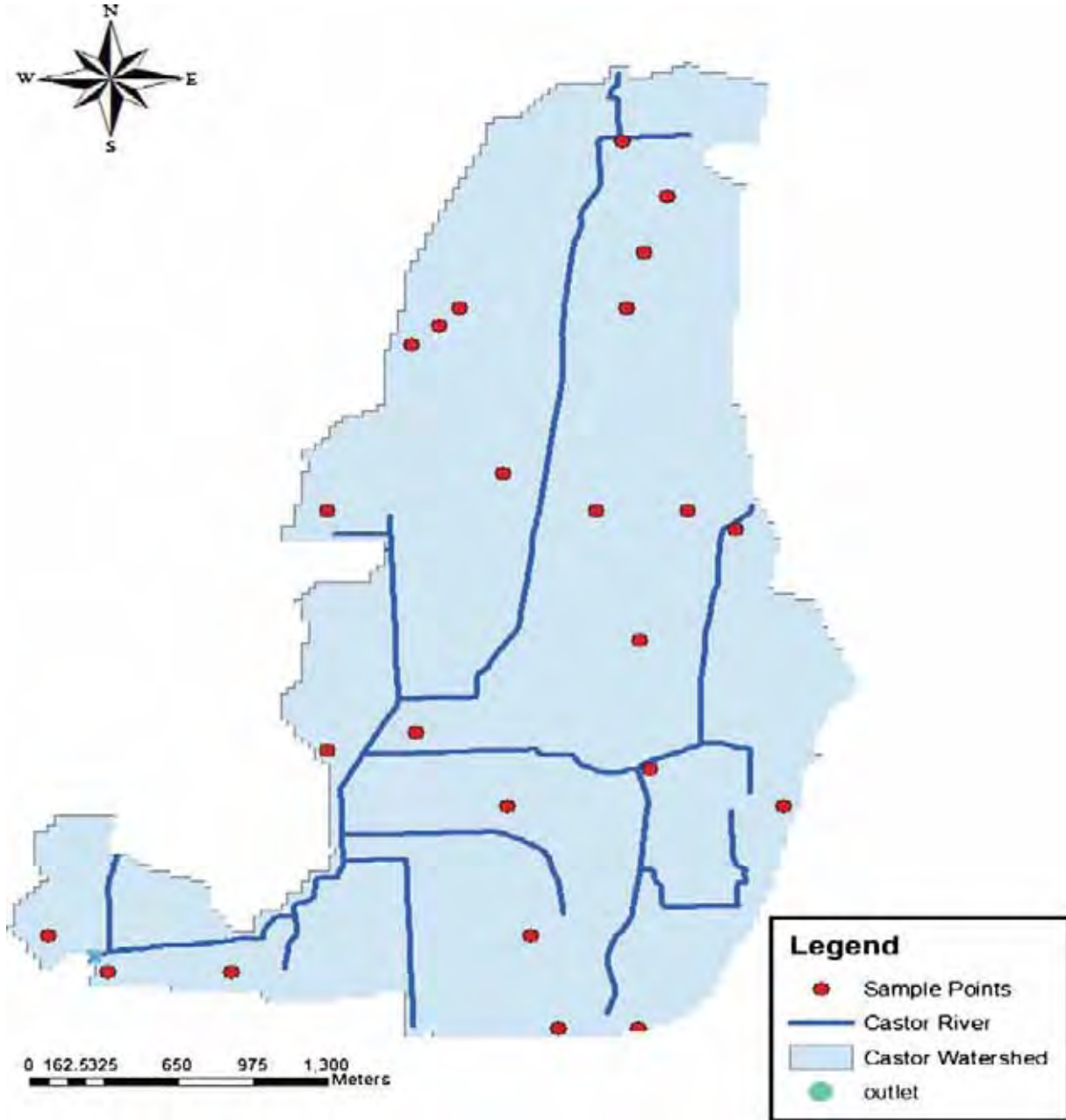
The major soil series of the watershed are as follows (Michaud 2004):

- Ste. Rosalie: poorly drained lacustrine and marine clays
- Bedford: an Orthic Humic Gleysol, fine loamy mixed calcareous
- Ste. Brigide: an Orthic Humic Gleysol, Coarse-loamy mixed calcareous
- St. Sebastien: Gleyed Sombric Brunisol, loamy-skeletal, mixed, monoacid, mesic

### 2.2 Field Measurements and Laboratory Analysis of Soil Properties

A stratified random soil sampling of the study area was also completed (Fig. 1). Soil samples (at depths 0–0.30, 0.30–0.60, and 0.60–0.90 m) were collected. SWAT soil database requires the following soil physical properties (Neitsch et al. 2009):

- *Soil bulk density (SOL\_BD)*. This quantity measures soil compaction. Among the factors that influence SOL\_BD include volumetric water content, hydraulic conductivity, soil water content, and soil porosity. The core samples were weighed and oven dried to constant weight at 105 °C. SOL\_BD was calculated as (Neitsch et al. 2009)



**Fig. 1** The study area and sample points

$$\rho = \frac{(M_W - M_D)}{V} \tag{1}$$

where  $\rho$  is the bulk density ( $\text{g cm}^{-3}$ ),  $M_W$  is the weight of wet soil (before oven dry) (g),  $M_D$  is the weight of dry soil (after oven dry) (g), and  $V$  is the volume of soil core in ( $\text{cm}^3$ ).

- *Particle size distribution (SOL\_CLAY, SOL\_SILT, and SOL\_SAND)*. Hydrometer method was used to analyze the percentage of clay, sand and silt (Neitsch et al. 2009; Somenahally et al. 2009). The unit is percentage of soil by weight.
- *Soil organic carbon (SOL\_CBN)*. Soil organic matter is related to soil stability, structure, fertility, and

nutrient retention. The loss-on-ignition procedure was used. This is based on the concept of using a very high temperature to convert the carbon to carbon dioxide. A soil mass of about 20 to 30 g was used for this process (Neitsch et al. 2009; Prasad and Power 1997; Donkin 1991). The unit is percentage of soil weight.

- *Available water content (SOL\_AWC)*. This is calculated as the difference between the fraction of water present at wilting point and that present at field capacity. A pressure plate device was used to estimate the soil field capacity and permanent wilting point. We determined the field capacity ( $\psi_{fc}$ ) as the amount of water content at a soil matric potential of  $-0.033$  bars. The permanent wilting point ( $\psi_{pwp}$ ) was determined as the water content at a soil matric potential of  $-15$  bars (Neitsch et al. 2009; Burk and Dalglish 2008). The unit is millimeter water per millimeter soil ( $\text{mm H}_2\text{O mm}^{-1}$  soil).
- *Universal soil loss equation's soil erodibility factor (K) (USLE\_K)*. This quantifies how one soil is more erodible than another when all other factors are the same (the unit is  $0.013 \text{ Mg m}^2 \text{ h} / \text{m}^3\text{-Mg cm}$ ). This can also be defined as the soil loss rate per erosion index unit for a specified soil as measured on a unit plot (Neitsch et al. 2009; Wischmeier and Smith 1978). The procedure suggested by (Neitsch et al. 2009) was used to derive the USLE\_K.
- *Hydraulic conductivity (SOL\_K) ( $\text{mm h}^{-1}$ )*. The mean soil texture in this study area is fine textured; therefore, the falling-head method was used to measure the soil permeability characteristics. The falling-head permeability test was used (Bowles 1986).

### 2.3 Spatial Clustering and Discretization of Soil Polygons

To account for spatial heterogeneity in soil properties, actual soil values instead of averages should be used. Therefore, clustering of the measured point set data was performed. This will provide the platform and the ability to define any desired number of spatial objects (soil polygons). This was done through the following procedures:

1. The Thiessen polygon tool in ArcGIS software 10 (Environmental Systems Research Institute, Redlands, CA, USA) was used to divide the area covered by each sample point (input feature) into Thiessen or proximal zones (Guo 2008). These

proximal polygons are very important in aggregation and division of sample points into desired heterogeneous regions. This process divided the study area into triangulated irregular networks that we considered heterogeneous. The concept of a Thiessen polygon is defined as a process where  $D$  is a set of points in coordinate or Euclidean space  $(x, y)$ . For any point  $k$  in that space, there is one point of  $D$  closest to  $k$ , except where point  $k$  is equidistant to two or more points of  $D$  (ESRI 2012).

2. As clustering these individual proximal zones was of interest, regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP) technique was implemented (Guo 2008; Bernassi and Ferrara 2010). Regionalization is very important in dividing spatial objects (proximal polygons) into a number of spatially contiguous regions. A full-order complete-linkage clustering (CLK) technique (Guo 2008) was used to derive heterogeneous clustered regions. This procedure involved two steps:

- 1) Clustering the data with contiguous constraints to produce spatially contiguous trees or hierarchy, and
- 2) Partitioning the tree to generate regions while optimizing an objective function

According to Guo (2008), the CLK defines the distance between two clusters as the dissimilarity between the farthest pair of points as

$$X_{CLK}(P, R) = \text{Max}_{a \in P, b \in R} (X_{a,b}), \quad (2)$$

where  $P$  and  $R$  are two clusters,  $a \in P$  and  $b \in R$  are the data points, and  $X_{a,b}$  is the dissimilarity between  $a$  and  $b$ .

Guo (2008) defined full-order constraining strategies as a clustering process where all edges are included in the process. The advantage of using full order over the first order is that the full-order strategy is dynamic and it updates the contiguity matrix after each merge (Guo 2008). This method produces different trees (Guo 2008), which therefore define different search methods for the partitioning. Thereafter, the partitioning of the contiguous tree will produce a number of sub-trees, each of which corresponds to a contiguous region (Bernassi and Ferrara 2010). The partitioning is iteratively done to partition a spatial contiguous tree into a number of regions by cutting a sub-tree (Guo 2008; Bernassi and Ferrara



2010). The following equation gives the homogeneity gain. This can also be called heterogeneity reduction. The best sub-tree to cut at each step is the sub-tree with the largest homogeneity expressed as

$$h_g^*(k) = \max[L(K_a) - L(K_b)] \quad (3)$$

where  $K_a$  and  $K_b$  are the two sub-trees from a possible cut of  $K$  and  $h_g^*$  is the homogeneity gain for the tree after the best cut (Guo 2008; Bernassi and Ferrara 2010).

The sum of square deviation (SSD), which can be used as the heterogeneity measure, is expressed as (Guo 2008)

$$L(K) = \sum_{h=1}^S \frac{s}{h} \sum_{g=1}^{n_r} \frac{n_r}{g} (x_{hg} - \gamma_h)^2 \quad (4)$$

where  $n_r$  is the number of objects in region  $K$ ,  $x_{gh}$  is the value for the  $h$ th attribute of the  $g$ th object,  $\gamma_h$  is the mean value of the  $h$ th attribute for all objects,  $K$  is a region,  $L(K)$  is the heterogeneity, and  $S$  is the number of attributes.

Therefore, the total heterogeneity,  $L_k$ , for a regionalization result with  $k$  regions can be given as the total of the  $k$  heterogeneity values (Guo 2008):

$$L_k = \sum_{h=1}^k L(K_h) \quad (5)$$

- From the foregoing, we partitioned the proximal zones (from (1) using the procedure explained in (2)) into 5, 10, 15, 20, and 24 heterogeneous clusters, termed "Region\_5", "Region\_10", "Region\_15", "Region\_20", and "Region\_24", respectively.
- We used the normal soil map with average properties that have been used in several studies (Deslandes et al. 2007; Michaud et al. 2008) in the watershed as the sixth configuration termed "Reference".

## 2.4 SWAT Model Setup

The SWAT model was selected for this project because of its worldwide applications (Gassman et al. 2007). SWAT2009 being the most stable recent version was selected along with an AVSWAT interface combined with ArcView 3.3.

SWAT requires several inputs such as a digital elevation model (DEM), land use, and soil characteristics.

Some climatic variables are also needed: daily rainfall, relative humidity, temperature, etc. All spatial inputs (DEM, soil (termed Reference later) and land use) were obtained from the Institut de Recherche et de Développement en Agroenvironnement. DEM has a spatial resolution of 30 m, accurate to approximately  $\pm 1.3$  m (Michaud et al. 2008). The climatological data was derived from the three closest weather stations (Philipsburg, Farnham, and Sutton) in the area (MDDEPQ 2003, 2005). Based on previous hydrologic modeling done in the area using SWAT (Michaud et al. 2008), a recommended value of  $31 \text{ mg kg}^{-1}$  was adopted for the initialized labile P soil concentration.

A sub-surface drainage of approximate 60 % for cultivated land, mean drain depth of 0.900 m, time required to achieve field capacity of 48 h, and drainage time into waterways of 18 h were assumed for the study area (Michaud et al. 2008).

Finally, we set up six SWAT model scenarios using the different soil configurations:

- Reference*. This was the SWAT model setup using soil map that has been used for previous studies in the study area. We call this reference because it is based on average values of soil properties.
- Region\_5*. This was a SWAT configuration setup using the survey soil properties partitioned into five heterogeneous regions.
- Region\_10*. This is a SWAT configuration setup using the survey soil properties partitioned into ten heterogeneous regions.
- Region\_15*. This is a SWAT configuration setup using the survey soil properties partitioned into 15 heterogeneous regions.
- Region\_20*. This is a SWAT configuration setup using the survey soil properties partitioned into 20 heterogeneous regions.
- Region\_24*. This is a SWAT configuration setup using the survey soil properties partitioned into 24 heterogeneous regions. In other words, in this configuration, each unit (sampled points) is a region in itself.

## 2.5 Evaluating Model

The model's output was obtained as a monthly time step. The model's performance was assessed using the coefficient of determination ( $R^2$ ) and the Nash–Sutcliffe efficiency (NSE) (Nash and Sutcliffe 1970). The  $R^2$  is a

statistical index which varies from 0 to 1, where  $R^2=1.0$  indicates that the correspondence between the predicted and observed perfectly fits a linear relationship. The NSE also ranges from  $-\infty$  to 1. An NSE of 1 indicates a perfect fit between the predicted and observed values, whereas an NSE of 0 represents predictions that are only as accurate as the mean of the observed data. When the NSE is less than zero, then the observed mean is a better predictor than the model. The simulation period was from January 1971 to December 2007. The periods from January 1971 to December 1976 (5 years) were used as a warm-up period. This is done to allow the model to “stabilize”, especially for model parameters such as soil moisture where they are dependent on prior conditions. The time periods from April 2001 to December 2002 were used for the evaluation of flow, sediment, and TP fluxes (before and after calibration). Observed discharges, suspended sediments, and TP data sets were obtained from the Ministère du Développement durable, Environnement et Parcs du Québec (MDDEPQ 2003, 2005).

### 3 Results and Discussion

#### 3.1 Descriptive Statistics of the Measured Soil Properties

Descriptive statistics of surveyed soil properties SOL\_BD, SOL\_CLAY, SOL\_SILT, SOL\_SAND, SOL\_CBN, SOL\_AWC, and SOL\_K are presented in Tables 1, 2, and 3 for soil depths of 0–0.30, 0.30–0.60, and 0.60–0.90 m, respectively. The parameter ranges are within the acceptable ranges required by SWAT (Neitsch et al.

2009). In examining the coefficient of variation for the three depths, SOL\_K and SOL\_CLAY were found to have the highest variability, while SOL\_BD has the lowest. This is consistent with findings from the literature concerning soil permeability and bulk density.

#### 3.2 Soil Clustering Analysis, Partitioning, and Regionalization

The result of the spatial division of the study area into 24 polygons using the Thiessen polygon technique in ArcGIS 10 is illustrated in Fig. 2. Each polygon has unique soil properties. We clustered these proximal zones into 24, 20, 15, 10, and 5 heterogeneous zones. The clusters and derived soil maps can be seen in Figs. 3, 4, and 5. Within-region heterogeneity of each of the regions is shown in Fig. 6. The SSD measure of within-region heterogeneity indicated that the configurations with the least number of regions (Region\_5) were the most heterogeneous while the lowest SSD value was found for Region\_24. In other words, the smallest SSD value was obtained when each unit (sample point) was a region in itself. This makes the region mean the same as the unit mean. Therefore, the greater number of regions, the less heterogeneous the region is. These maps were used as soil inputs into SWAT to quantify the impact of heterogeneities of the measured soil properties.

#### 3.3 Impacts of Soil Heterogeneities

##### 3.3.1 Hydrologic Response Units

The HRU is a unique combination of land use pattern, soil types, and landscape attributes. The different HRUs for

**Table 1** Descriptive statistics of surveyed soil properties (0–0.30 m)

	SOL_BD1	SOL_AWC1	SOL_K1	SOL_CBN1	CLAY1	SILT1	SAND1	USLE_K1
Mean	1.34	0.061	60.90	4.28	18.18	31.42	50.39	0.23
Standard error	0.04	0.005	9.02	0.23	3.06	1.9077	4.1070	0.029
Median	1.34	0.054	49.14	4.15	12.70	29.43	56.46	0.21
Standard deviation	0.20	0.028	44.19	1.15	15.029	9.346	20.120	0.14
Sample variance	0.040	0.00080	1953.18	1.33	225.89	87.35	404.82	0.020
Kurtosis	0.42	−0.086	−1.62	−0.028	−0.4602	−0.50	−0.993	−0.49
Skewness	0.27	0.95	0.36	0.502	0.902	0.706	−0.599	0.343
Range	0.85	0.091	110.29	4.59	48.79	32.640	63.046	0.508
Minimum	0.98	0.029	10.11	2.37	2.61	18.186	10.972	0.0085
Maximum	1.84	0.12	120.40	6.96	51.41	50.82	74.019	0.51

**Table 2** Descriptive statistics of surveyed soil properties (0.30–0.60 m)

	SOL_BD2	SOL_AWC2	SOL_K2	SOL_CBN2	CLAY2	SILT2	SAND2	USLE_K2
Mean	1.40	0.068	35.52	3.503	26.29	35.78	37.91	0.26
Standard error	0.03	0.006	7.35	0.29	4.15	2.21	4.47	0.037
Median	1.35	0.058	26.43	3.139	23.16	33.18	31.36	0.22
Standard deviation	0.16	0.029	36.02	1.43	20.36	10.87	21.94	0.18
Sample variance	0.026	0.00087	1297.60	2.06	414.55	118.17	481.48	0.033
Kurtosis	3.48	0.171	-1.31	1.44	-1.26	2.29	-1.48	2.35
Skewness	2.00	0.99	0.70	0.33	0.47	1.38	0.074	1.29
Range	0.64	0.104	87.38	6.80	60.40	44.08	66.30	0.80
Minimum	1.24	0.032	0.88	0	5.08	20.94	3.085	0.004
Maximum	1.89	0.13	88.26	6.80	65.48	65.020	69.38	0.811

each soil configuration are shown in Fig. 7. There is a significant difference among the various configurations. This makes sense since the soils with higher region numbers are expected to have more soil classes and therefore higher HRUs. Since HRU is where the simulation of runoff, sediment, and nutrients starts, we should expect a proportionate increase in the magnitude of the simulations.

### 3.3.2 Before Calibration

The following factors are considered before calibration:

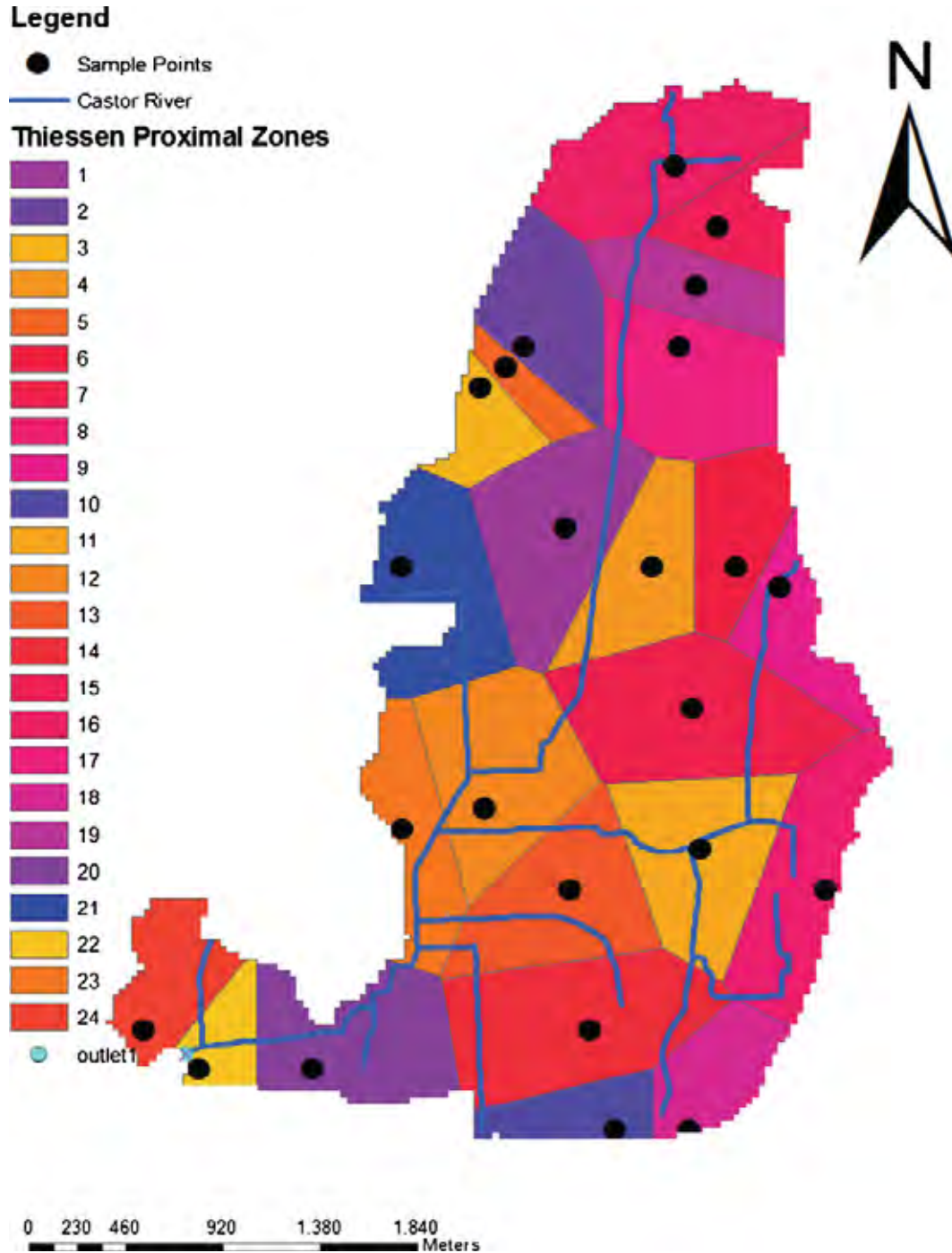
- *Runoff*. Graphical comparisons between the model simulation and observation of surface flow between April 2001 and December 2002 are shown in Fig. 8. For the five different configurations, no significant differences are seen (Fig. 8). This may be attributable to the SCS–curve number (CN) method implemented in SWAT. Surface runoff is estimated using the SCS

runoff equation. This is an empirical model and could contribute to the variability of surface runoff. Ye et al. (2009) found that the SCS–CN method weakens the discrepancy between the different resolutions of soil heterogeneities and strongly affects the similarity in flow prediction. In other words, the CN threshold determined by the soil hydrologic groups are ranked based on soil permeability (Ye et al. 2009; Mishra and Singh 2003). Using this group across all soil types often masked out soils that have notable differences in physical characteristics.

The NSE values for the prediction of surface runoff were 0.60, 0.57, 0.60, 0.59, 0.59, and 0.41, and the  $R^2$  values were 0.72, 0.67, 0.69, 0.69, and 0.63, for the Reference, Region\_5, Region\_10, Region\_15, Region\_20, and Region\_24, respectively (Table 4). The performance evaluation was quite good for an uncalibrated model. In other words, to increase these indices, we need to adjust some parameters as would

**Table 3** Descriptive statistics of surveyed soil properties (0.60–0.90 m)

	SOL_BD3	SOL_AWC3	SOL_K3	SOL_CBN3	CLAY3	SILT3	SAND3	USLE_K3
Mean	1.21	0.073	1.90	2.62	31.23	30.35	25.91	0.27
Standard error	0.099	0.0067	0.31	0.31	5.51	3.90	4.93	0.048
Median	1.39	0.070	1.25	2.54	21.70	29.08	18.50	0.22
Standard deviation	0.48	0.033	1.52	1.54	27.02	19.15	24.16	0.23
Sample variance	0.23	0.001	2.31	2.37	730.25	366.78	583.73	0.05
Kurtosis	3.23	0.71	-1.24	1.12	-1.156	0.971	-1.23	0.31
Skewness	-2.09	1.07	0.41	0.38	0.56	0.61	0.57	0.86
Range	1.69	0.11	4.40	6.61	76.71	79.84	66.91	0.86
Minimum	0	0.033	0	0	0	0	0	0
Maximum	1.69	0.14	4.40	6.61	76.71	79.8	66.91	0.86



**Fig. 2** Thiessen polygons or proximal zones derived from sample points using ArcGIS 10 (Note 1–24 are the geo-referenced points where soil samples were collected)

be done under sensitivity analysis. The significant underestimation, especially in the period of January–March 2002, has been attributed to the exceptionally unseasonable near-zero and above-zero temperatures

(some as high as 10 °C) recorded. SWAT had difficulty in distinguishing between rainfall and snowfall given its weather-generating method (Deslandes et al. 2004; Michaud et al. 2007)



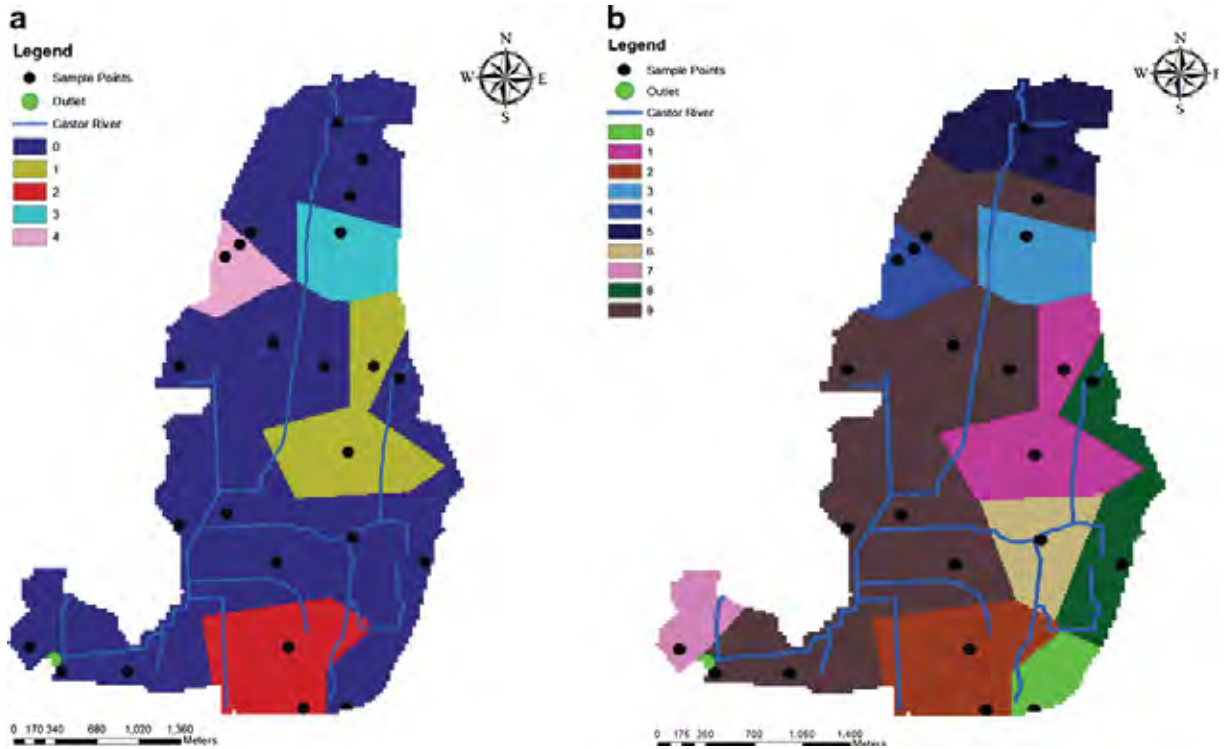


Fig. 3 Heterogeneous soil clusters derived using REDCAP a five and b ten regions

- *Sediment.* The time series plot of the simulated and observed sediment load is shown in Fig. 9. A

comparison shows that SWAT configurations capture the peak flow periods and also the low flow periods.

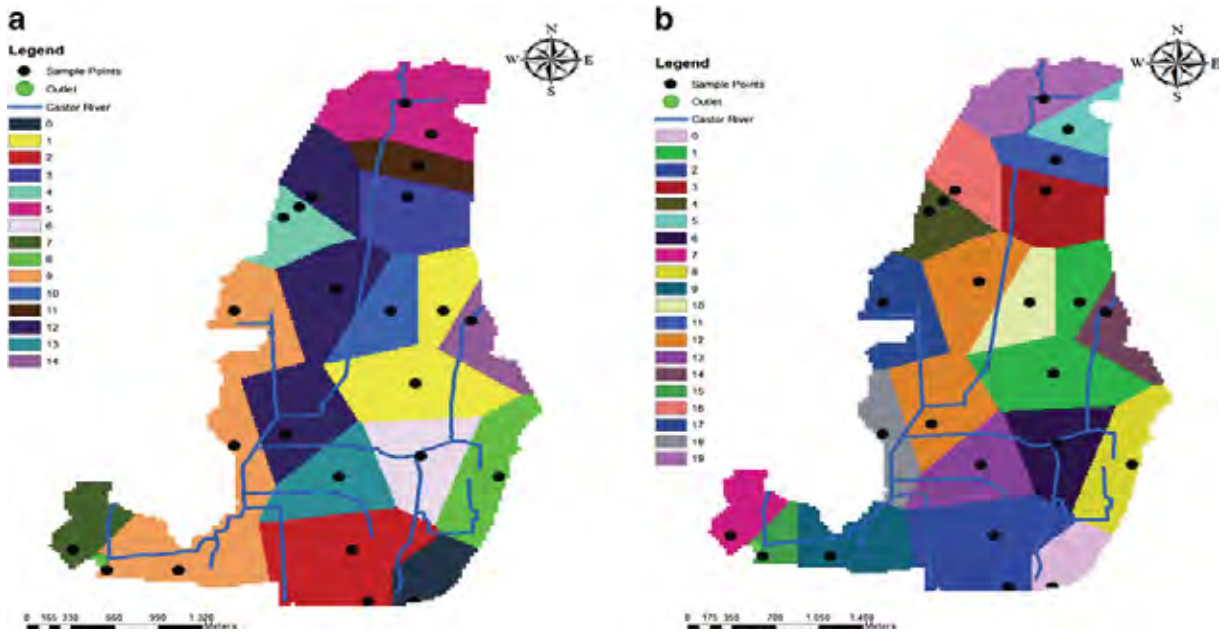


Fig. 4 Heterogeneous soil clusters derived using REDCAP a 15 and b 20 regions



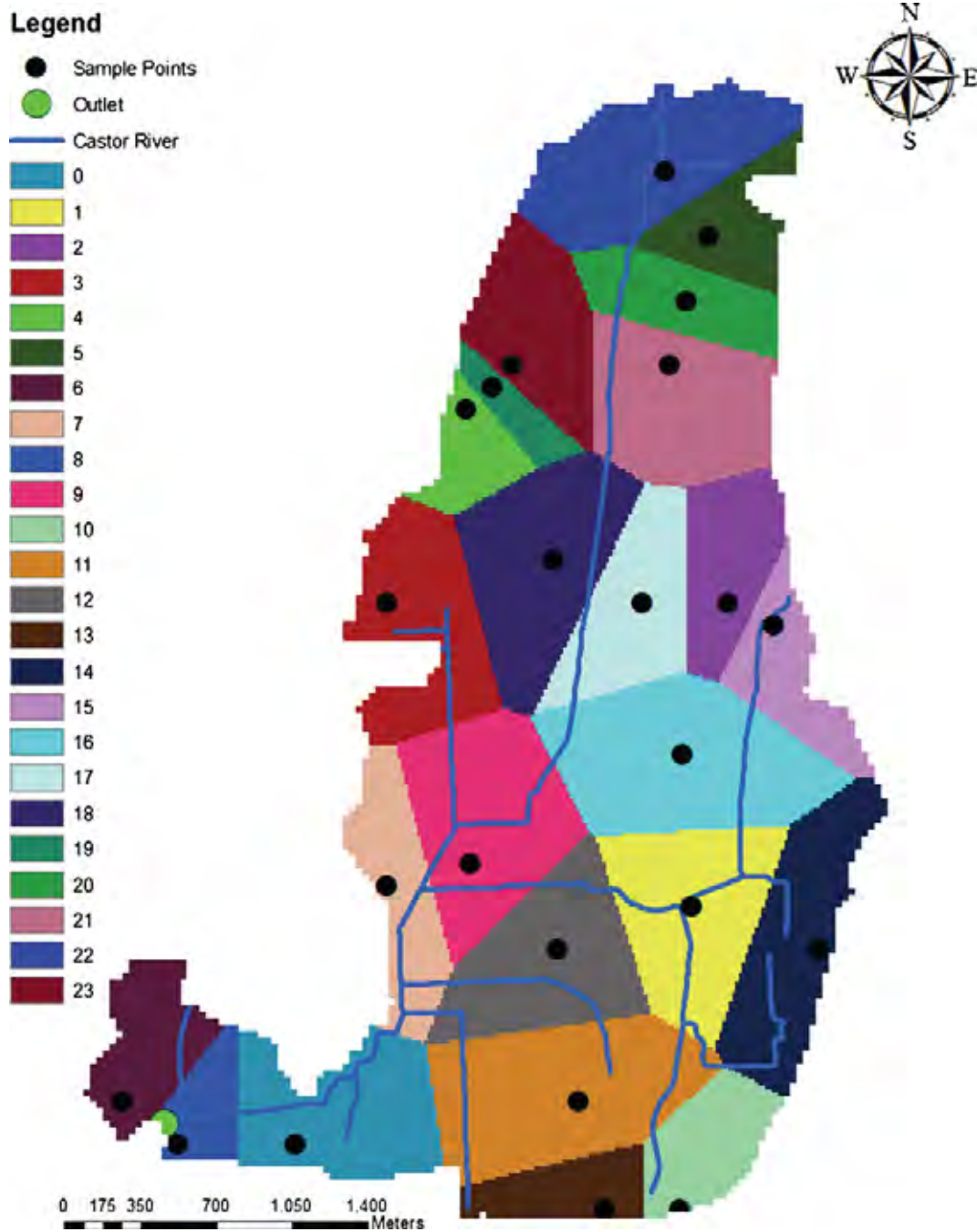
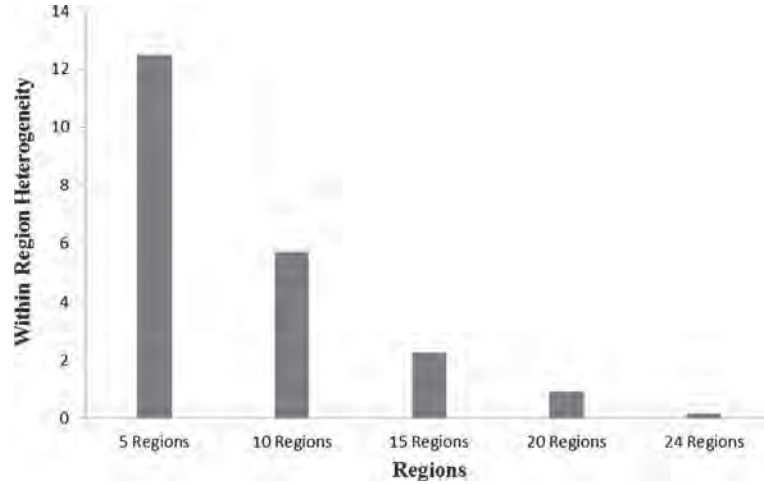


Fig. 5 Twenty-four heterogeneous soil clusters derived using REDCAP

The NSE was poor, being less than 0.3 in all cases (Table 4). The  $R^2$  was 0.58 in all cases, except for Region\_24 for which it was 0.45 (Table 4).

- *Total phosphorus*. The plot of measured vs. uncalibrated TP load predicted is shown in Fig. 10. Although the SWAT configurations captured high

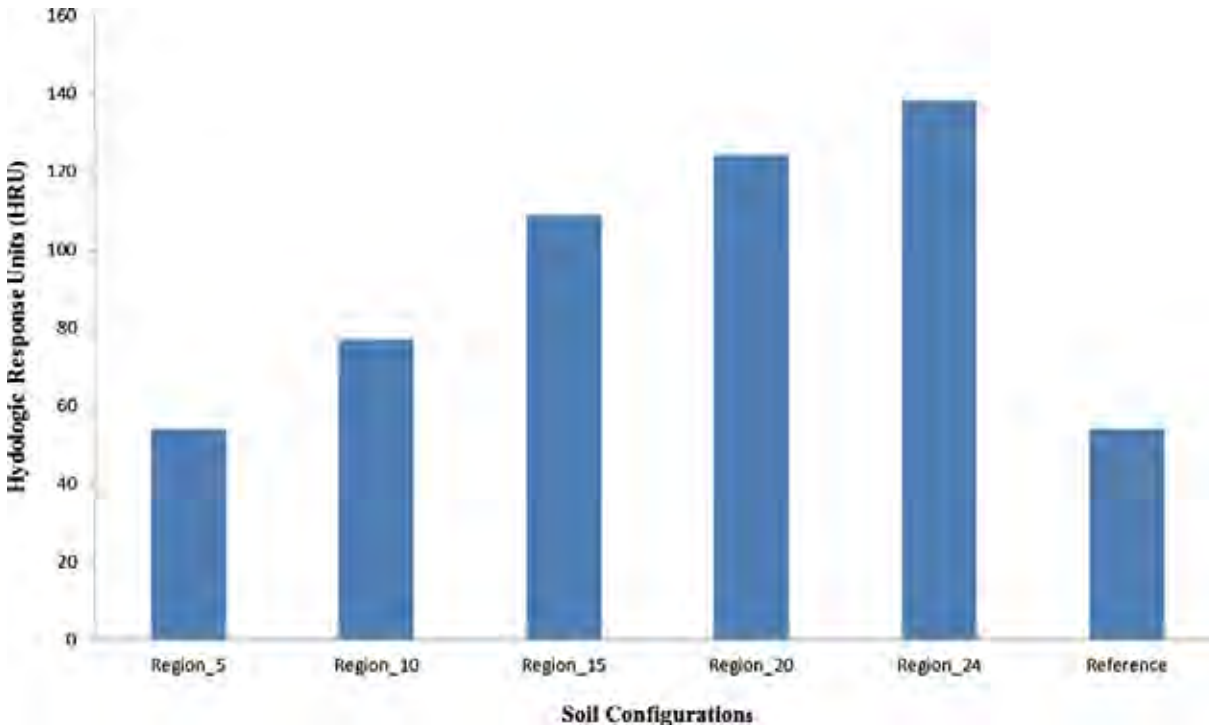
**Fig. 6** Within-region measured heterogeneity



fluxes due to high flows, in all cases, it underpredicted the P fluxes. The coefficients of determination were 0.83, 0.78, 0.85, 0.85, and 0.88 for Reference, Region\_5, Region\_10, Region\_15, Region\_20, and Region\_24, respectively (Table 4).

### 3.3.3 Sensitivity Analysis

Before calibration, sensitivity analysis is a process that must be performed to know which parameters to adjust during the calibration procedure and by how much. We



**Fig. 7** HRUs across all configurations

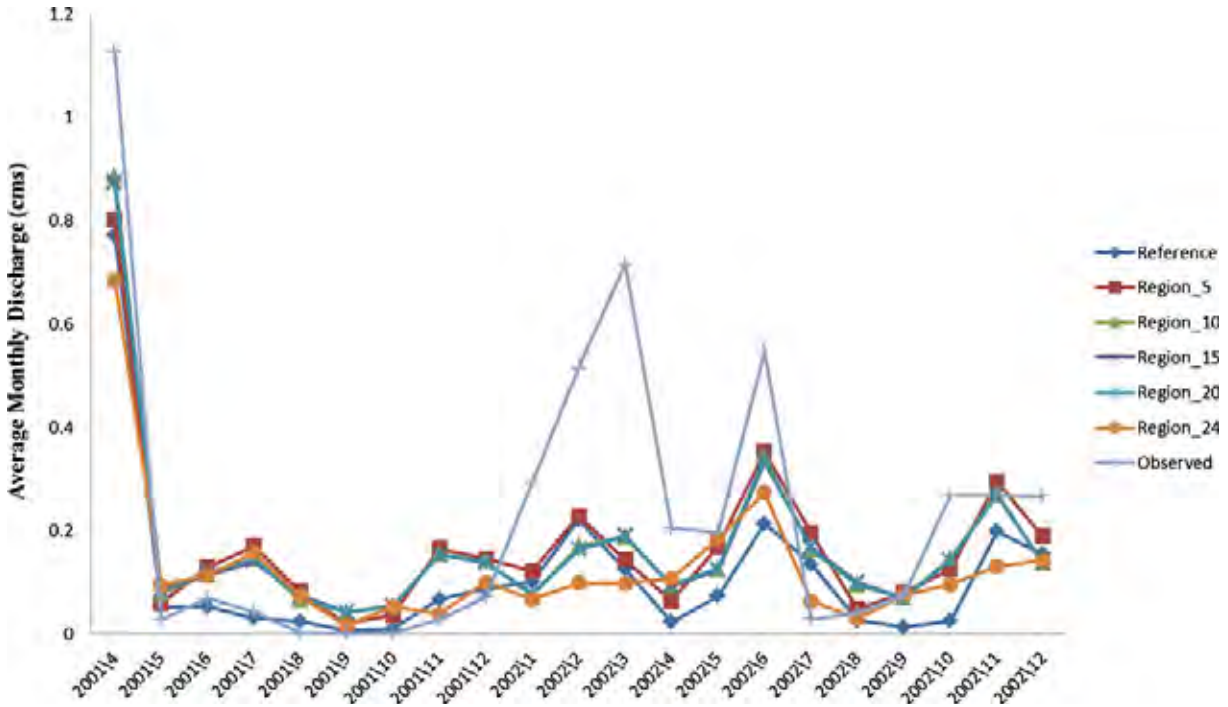


Fig. 8 Plot of the comparison between averages of monthly simulated and observed discharge (April 2001–December 2002)

used the sequential uncertainty fitting (SUFI) version 2 algorithm in SWAT calibration and uncertainty procedures to carry out this process. SUFI is used to derive the posterior parameters from priors. In other words, parameter uncertainty accounts for all sources of uncertainties

such as the uncertainties in inputs (such as rainfall), model structure, and observed data (Abbaspour et al. 2007). A 95 % uncertainty is obtained at the 2.5 and 97.5 % levels of the cumulative distribution of the input variable derived by the Latin hypercube sampling (Abbaspour et al.

Table 4 The Performance assessment of SWAT configurations using  $R^2$  and the Nash–Sutcliffe coefficient

Model	Reference		Region									
	$R^2$	NSE	5		10		15		20		24	
			$R^2$	NSE	$R^2$	NSE	$R^2$	NSE	$R^2$	NSE	$R^2$	NSE
Monthly discharge (cm)												
Non-calibrated	0.72	0.60	0.67	0.57	0.69	0.60	0.69	0.59	0.69	0.59	0.63	0.41
Calibrated	0.65	0.54	0.64	0.59	0.65	0.59	0.65	0.59	0.65	0.59	0.60	0.48
Sediment load (Mg)												
Non-calibrated	0.58	-0.56	0.57	-0.64	0.58	-0.21	0.58	-0.18	0.58	-0.10	0.45	-2.94
Calibrated	0.50	0.21	0.50	0.23	0.50	0.21	0.50	0.23	0.47	0.27	0.40	0.10
Total phosphorus (kg)												
Non-calibrated	0.83	0.51	0.78	0.53	0.85	0.32	0.85	0.48	0.85	0.52	0.88	0.28
Calibrated	0.81	0.45	0.71	0.51	0.80	0.45	0.86	0.55	0.81	0.56	0.80	0.45

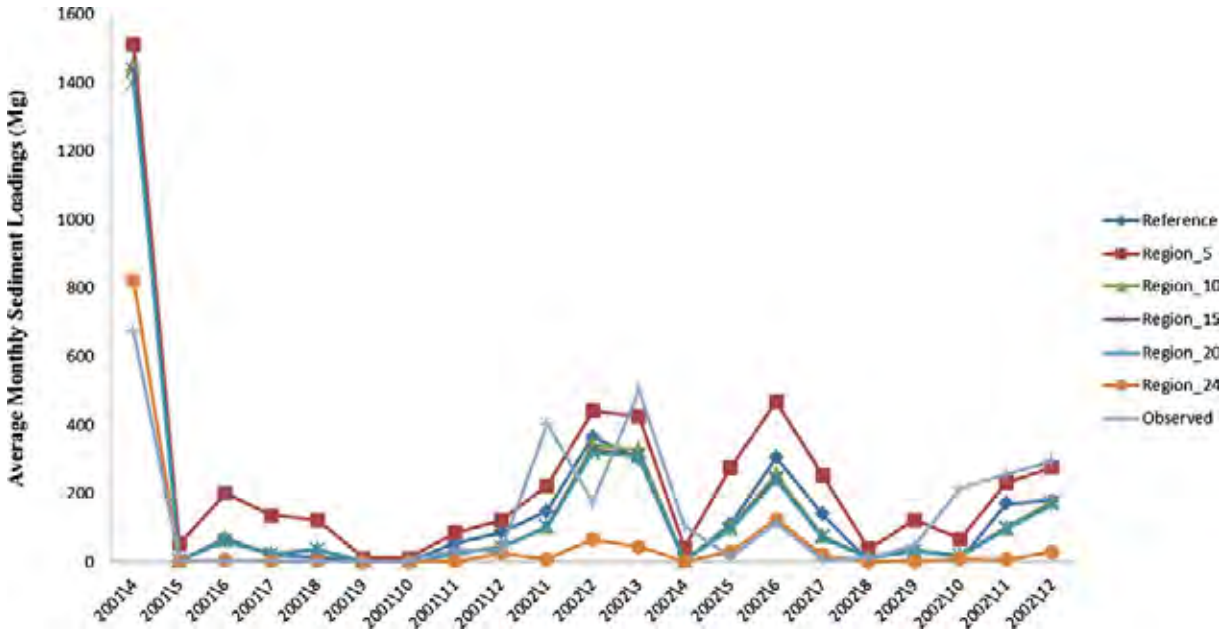


Fig. 9 Plot of the comparison between averages of monthly simulated and observed sediment load (April 2001–December 2002)

2007). The parameters shown in Table 5 were found to be the most sensitive for hydrology for all SWAT configurations after 1,000 simulations. The *t*-stat provides a measure of sensitivity, while the *p* values determined

the significance of the sensitivity (Abbaspour et al. 2007). A value close to zero has more significance (Abbaspour et al. 2007). Therefore, in Table 5, the base flow factor (ALFA\_BF), base flow alpha factor for bank

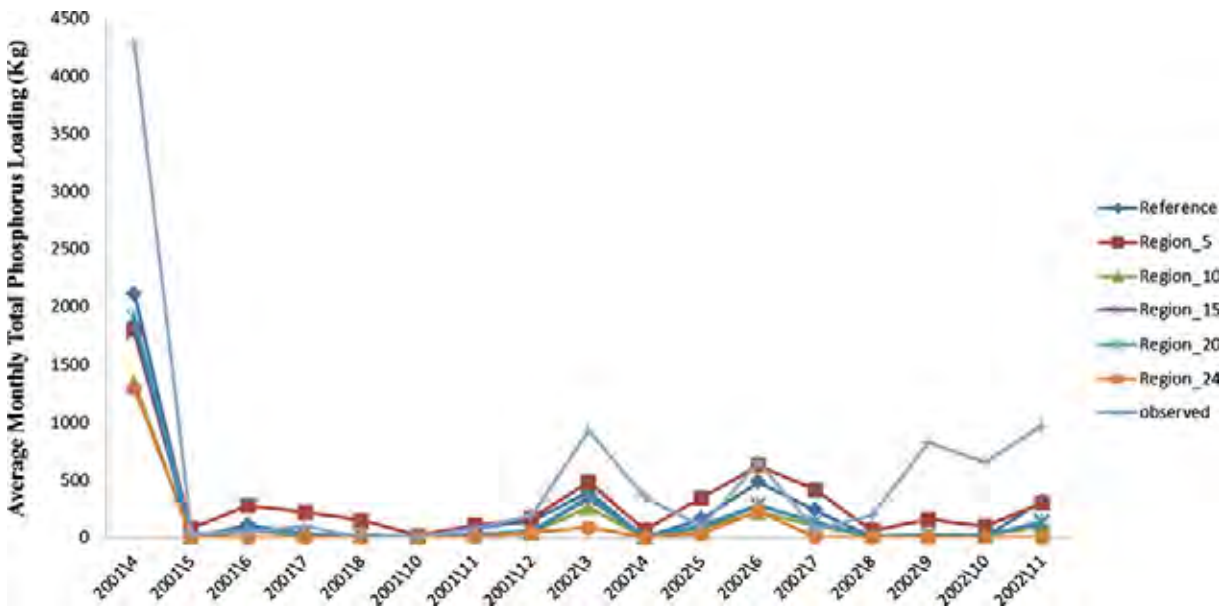


Fig. 10 Plot of the comparison between averages of monthly simulated and observed total phosphorus load (April 2001–November 2002)

**Table 5** The sensitivity analysis results for SWAT configurations

Parameter name	<i>t</i> -Stat	<i>P</i> value
<i>v</i> __GW_REVAP.gw <sup>a</sup>	0.08	0.93
<i>v</i> __GW_DELAY.gw <sup>a</sup>	-0.10	0.92
<i>r</i> __SOL_AWC(1).sol <sup>b</sup>	-0.14	0.89
<i>r</i> __SOL_K(1).sol <sup>b</sup>	-0.21	0.83
<i>v</i> __CH_K2.rte <sup>a</sup>	-0.31	0.75
<i>v</i> __ESCO.hru <sup>a</sup>	-0.44	0.66
<i>v</i> __SFTMP.bsn <sup>a</sup>	0.68	0.50
<i>r</i> __SOL_BD(1).sol <sup>b</sup>	-0.68	0.50
<i>r</i> __CN2.mgt <sup>b</sup>	-0.69	0.49
<i>v</i> __GWQMN.gw <sup>a</sup>	0.73	0.46
<i>v</i> __CH_N2.rte <sup>a</sup>	1.70	0.09
<i>v</i> __ALPHA_BNK.rte <sup>a</sup>	3.09	0.00
<i>v</i> __ALPHA_BF.gw <sup>a</sup>	22.41	0.00

*GW\_REVAP* groundwater “revap” coefficient, *GW\_DELAY* groundwater delay (days), *SOL\_AWC* available water capacity of the soil layer, *SOL\_K* saturated hydraulic conductivity, *ESCO* soil evaporation compensation factor, *SFTMP* snowfall temperature, *SOL\_BD* moist bulk density, *CN2* SCS runoff curve number, *GWQMN* threshold depth of water in the shallow aquifer required for return flow to occur (mm), *CH\_N2* Manning’s “n” value for the main channel, *CH\_K2* effective hydraulic conductivity in main channel alluvium, *ALPHA\_BNK* base flow alpha factor for bank storage, *ALPHA\_BF* base flow alpha factor (days)

<sup>a</sup> *v*\_\_ means the existing parameter value is to be replaced by the given value

<sup>b</sup> *r*\_\_ means the existing parameter value is multiplied by 1+ a given value

storage (*ALPHA\_BNK*), and Manning’s “n” value for the main channel (*CH\_N2*) are the most sensitive parameters followed by SCS runoff *CN2*.

### 3.3.4 After Calibration for Hydrology, Sediment, and TP

In terms of hydrology, model performance was evaluated using  $R^2$  and NSE. In all cases,  $R^2$  is greater than 0.6, while NSE is up to 0.6 for all configurations except Region\_24 (Table 4). As for sediment load, USLE support, biological mixing efficiency, peak rate adjustment factor for sediment routing in the main channel, and peak rate adjustment factor for sediment routing in the sub-basins were the parameters that were adjusted. There were significant improvements in the models’ performance based on NSE. The  $R^2$  also increases. For TP, SWAT predictions were hardly improved compared to the uncalibrated results. The following parameters were adjusted, namely phosphorus

sorption coefficient, phosphorus percolation coefficient, and phosphorus soil partitioning coefficient.  $R^2$  equally ranges from 0.72 to 0.85 across all configurations.

Generally, the essence of calibration in hydrology is to modify the input parameters to a hydrologic model until the output from the model matches an observed set of data. There are uncertainties in the observed set of data. Also, the conceptualization or structure of the model is not free from uncertainties. Therefore, it is possible to have lower performance indices such as  $R^2$  and NSE after calibration as seen in this study and in other literature (Srinivasan et al. 2010). Also, Mukundan et al. (2010) reported that calibration can mask important parametric information. In fact, Heathman et al. (2008) attested that hydrologic models in uncalibrated mode eliminate bias due to parameter optimization.

Overall, it can be said that all SWAT configurations perform at the same scale compared with the observed data. Putting the reference configuration into perspective, one may not necessarily be losing much information using the average values. In other words, having detailed or high-resolution information, especially for soil properties, does not necessarily improve the prediction in the face of model structural uncertainty. Use of CN in SWAT tends to mask or cover what would have been the impact of different soil properties.

## 4 Concluding Remarks

The impact of soil heterogeneity in predicting runoff, sediment load, and TP was examined in this study. Five different soil configurations that were partitioned and regionalized to account for heterogeneity at various stages (5, 10, 15, 20, and 24 heterogeneous regions) were investigated. This method showed that when a sample point is a region unto itself, the measure of within-region heterogeneity is lowest. In other words, the greater the number of regions, the lesser the within-region heterogeneity is. From an HRU perspective, there are significant differences in their distributions. However, this does not translate into differences in model prediction.

Results showed that an increase in spatial detail by taking more soil measurements does not necessarily increase the accuracy of the model predictions. The use of curve numbers in SWAT tends to mask any discrepancies because it is based on the soil hydrologic group, which conceals the different physical properties of the soil. This is part of the uncertainty in the model structure that needs



to be addressed. Therefore, hydrologists need to weigh the pros and cons of having high-resolution data and the cost involved. As demonstrated in this study, having more samples and using them as individual units without averaging does not necessarily improve the predictions.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

- Abbaspour, K. C., Vajdani, M., & Haghighat, S. (2007). SWAT-564 CUP calibration and uncertainty programs for SWAT. [http://www.mssanz.org.au/MODSIM07/papers/24\\_s17/SWAT-UP\\_s17\\_Abbaspour.pdf](http://www.mssanz.org.au/MODSIM07/papers/24_s17/SWAT-UP_s17_Abbaspour.pdf). Accessed 02 Feb 2013.
- Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologic modeling and assessment—Part I: model development. *Journal of the American Water Resources Association*, 34(1), 73–89.
- Bernassi, F., & Ferrara, R. (2010). *Regionalization with dynamically constrained agglomerative clustering and partitioning: an application on spatial segregation of foreign population in Italy at regional level*. In 45th Scientific meeting of the Italian statistical society. Padua: University of Padua. 29 July 2010b.
- Bowles, J. E. (1986). *Engineering properties of soils and their measurement*. New York: McGraw-Hill.
- Burk, L., & Dalglish, N. (2008). *Estimating plant available water capacity—a methodology*. Canberra: CSIRO Sustainable Ecosystems.
- Deslandes, J., Michaud, A., & Bonn, F. (2004). Use of GIS and remote sensing to develop indicators of phosphorus non-point source pollution in the Pike River Basin. In T. O. Manley, P. L. Manley, & T. B. Mihuc (Eds.), *Lake Champlain: partnerships and research in the new millennium* (pp. 271–290). New York: Kluwer Academic/Plenum Pub.
- Deslandes, J., Beaudin, I., Michaud, A., Bonn, F., & Madramootoo, C. A. (2007). Influence of landscape and cropping system on phosphorus mobility within the Pike River watershed of southwestern Quebec: model parameterization and validation. *Canadian Water Resources Journal*, 32(1), 21–42. doi:10.4296/cwrj3201021.
- Donkin, M. J. (1991). Loss-on-ignition as an estimator of soil organic carbon in A-horizons of forestry soils. *Communication in Soil Science Plant Analysis*, 22, 233–241.
- ESRI, (2012). Create Thiessen polygons. <http://resources.arcgis.com/en/help/main/10.1/index.html#//00080000001m000000>. Accessed 20 December 2012.
- Gassman, P. W., Reyes, M. R., Green, C. H., & Arnold, J. G. (2007). The soil and water assessment tool: historical development, applications and future research directions. *Transactions of the ASABE*, 50(4), 1211–1250.
- Guo, D. (2008). Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP). *International Journal of Geographical Information Science*, 22(7), 801–823.
- Heathman, G. C., Flanagan, D. C., Larose, M., & Zuercher, B. W. (2008). Application of the soil and water assessment tool and annualized agricultural non-point source models in the St. Joseph River watershed. *Journal of Soil and Water Conservation*. 63(6). doi:10.2489/jswc.63.6.552.
- Hegman, W., Wang, D., & Borer, C. (1999). *Estimation of Lake Champlain basinwide nonpoint source phosphorus export. Lake Champlain basin program technical report no. 31*. Grand Isle: LCBP.
- Jamieson, A., Madramootoo, C. A., & Enright, P. (2003). Phosphorus losses in surface and subsurface runoff from a snowmelt event on an agricultural field in Quebec. *Canadian Biosystems Engineering*, 45(1), 1–7.
- MDDEPQ (Ministère du Développement durable, Environnement et Parcs du Québec MDDEPQ). (2005). Le réseau hydro-métrique québécois 1998 à 2003. [Quebec's hydrological network 1998–2003.] MDDEPQ, Centre d'expertise hydrique du Québec. <http://www.cehq.gouv.qc.ca/hydrometrie/reseau/index.htm>. Accessed 02 Jan 2013.
- MDDEPQ (Ministère du Développement durable, Environnement et Parcs du Québec). (2003). Base de données Climatologique du Québec de 1997 à 2003. [Climatological database for Quebec, 1997–2003]. Service de l'information sur le milieu atmosphérique, Direction du suivi de l'état de l'environnement. [http://www.hc-sc.gc.ca/ewh-semt/pubs/eval/inventory-repertoire/climatologie\\_f.html](http://www.hc-sc.gc.ca/ewh-semt/pubs/eval/inventory-repertoire/climatologie_f.html). Accessed 02 Feb 2013.
- Medalie, L., & Smeltzer, E. (2004). Status and trends of phosphorus in Lake Champlain and its tributaries, 1990–2000. In T. O. Manley, P. L. Manley, & T. B. Mihuc (Eds.), *Lake Champlain: partnerships and research in the new millennium* (pp. 191–219). New York: Kluwer Academic/Plenum.
- Michaud, A. R. (2004). *Indicateurs Agroenvironnementaux Adaptes A la Gestion De Project Cibles Sur la Prévention De la Pollution Diffuse par le Phosphore. Faculté Des Sciences De l'agriculture et De l'alimentation*. Québec: Université Laval.
- Michaud, A. R., Beaudin, I., Deslandes, J., Bonn, F., & Madramootoo, C. A. (2007). SWAT-predicted influence of different landscape and cropping system alterations on phosphorus mobility within the pike river watershed of southwestern Québec. *Canadian Journal of Soil Science*, 87(3), 329–344.
- Michaud et al. (2008). Beneficial management practices and water quality: hydrological modeling of two basins in the Monteregion region (Quebec). National Agric-Environmental Standards Initiative Technical Series Report No. 4–65. Gatineau: Prepared and published by Environment Canada
- Mishra, S. K., & Singh, V. P. (2003). *Soil conservation service curve number (SCS-CN) methodology*. Dordrecht: Kluwer Academic Publishers.
- Mukundan, R., Radcliffe, D. E., & Risse, L. M. (2010). Spatial resolution of soil data and channel erosion effects on SWAT model predictions of flow and sediment. *Journal of Soil and Water Conservation*. doi:10.2489/jswc.65.2.92.
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models. Part I: a discussion of principles. *Journal of Hydrology*, 10(3), 282–290.
- Neitsch, S. L., Arnold, J. G., Kiniry, J. R., William, J. R., & King, K. W.; Grassland, Soil and Water Research Laboratory of the USDA Agricultural Research Service; Blackland Research

- and Extension Center (2009). Soil and water assessment tool theoretical documentation. College Station: Texas Water Resources Institute Report TR-191. <http://twri.tamu.edu/reports/2011/tr406.pdf>. Accessed 02 Feb 2013.
- Prasad, R., & Power, F. J. (1997). *Soil fertility management for sustainable agriculture*. New York: Lewis Publishers.
- Somenahally, A., Weindorf, D. C., Darilek, L., Muir, J. P., Wittie, R., Thompson, C., et al. (2009). Spatial variability of soil test phosphorus in manure-amended soils on three dairy farms in North Central Texas. *Journal of Soil and Water Conservation*, 64(2), 89–97.
- Srinivasan, R., Zhang, X., & Arnold, J. (2010). SWAT ungauged: hydrological budget and crop yield predictions in the Upper Mississippi River Basin. *Transactions of the ASABE*, 53(5), 1533–1546.
- Steinman, A. D., & Denning, R. (2005). The role of spatial heterogeneity in the management of freshwater resource. In G. M. Lovett, C. G. Jones, M. G. Turner, & K. C. Weathers (Eds.), *Ecosystem function in heterogeneous landscapes* (pp. 367–387). New York: Springer.
- Tague, C. (2005). Heterogeneity in hydrologic processes: a terrestrial hydrologic modeling perspective. In G. M. Lovett, C. G. Jones, M. G. Turner, & K. C. Weathers (Eds.), *Ecosystem function in heterogeneous landscapes* (pp. 119–135). New York: Springer.
- Wischmeier, W. H., & Smith, D. D. (1978). Predicting rainfall erosion losses—a guide to conservation planning. Washington, DC: U.S. Department of Agriculture (handbook no. 537).
- Ye, X. C., Viney, N. R., & Zhang, Q. (2009). Effects of spatial resolution of soil data on hydrological processes modelling: a case study using the SWAT model. 18th World IMACS/MODSIM Congress, Cairns, Australia 13–17 July, 2009.

# Soil bacteria and archaea found in long-term corn (*Zea mays* L.) agroecosystems in Quebec, Canada

Sara Sheibani<sup>1</sup>, Sandra F. Yanni<sup>1</sup>, Roland Wilhelm<sup>1</sup>, Joann K. Whalen<sup>1,4</sup>, Lyle G. Whyte<sup>1</sup>, Charles W. Greer<sup>1,2</sup>, and Chandra A. Madramootoo<sup>3</sup>

<sup>1</sup>Department of Natural Resource Sciences, Macdonald Campus of McGill University, Ste-Anne-de-Bellevue, Québec, Canada H9X 3V9; <sup>2</sup>National Research Council, Biological Research Institute, 6100 Royalmount Avenue, Montreal, Quebec, Canada; and <sup>3</sup>Department of Bioresource Engineering, Macdonald Campus, McGill University, 21,111 Lakeshore Road, Ste-Anne-de-Bellevue, Quebec, Canada.

Received 12 April 2012, accepted 11 October 2012.

Sheibani, S., Yanni, S. F., Wilhelm, R., Whalen, J. K., Whyte, L. G., Greer, C. W. and Madramootoo, C. A. 2013. **Soil bacteria and archaea found in long-term corn (*Zea mays* L.) agroecosystems in Quebec, Canada.** *Can. J. Soil Sci.* **93**: 45–57. The soil microbial community controls all biological processes in soils and is considered a good indicator of general soil health. Assessment of the microbial community in intensively cropped soils that are under reduced tillage management is especially important because the microbes are the primary decomposers of the high residue input in such systems. We investigated the microbial biomass and diversity of bacteria and archaea in a sandy-loam Dystric Gleysol from a long-term (15 yr) corn (*Zea mays* L.) agroecosystem in Quebec, Canada, under conventional (CT), reduced tillage (RT), and no tillage (NT) and two residue inputs (high level: +R and low level: –R). Analysis included microbial biomass C and N (MBC, MBN), catalyzed reporter deposition-fluorescence in situ hybridization (CARD-FISH) and 5-(4, 6-dichlorotriazinyl) amino fluorescein hydrochloride (DTAF) cell counts, 16S rRNA polymerase chain reaction-denaturing gradient gel electrophoresis (PCR-DGGE) and an archaeal clone library. The PCR-DGGE analysis identified Proteobacteria, Actinobacteria and Firmicutes as dominant groups in all tillage and residue management treatments. The archaeal group was diverse, with most individuals identified as belonging to the Crenarchaeota phylum. We also detected soil archaea belonging to the newly proposed phylum Thaumarchaeota, the chemolithoautotrophic ammonia-oxidizing archaeota, in a corn agroecosystem in Quebec, Canada. Microbial biomass increased in the +R treatment according to MBC concentration and direct cell counts. Considering results from the CARD-FISH counts (bacterial and archaeal cell counts without fungal cells) and from MBC results (all microbial biomass including fungi) we concluded the likelihood of greater fungal biomass in the NT plots.

**Key words:** Archaeal phylogeny, conservation tillage, corn agroecosystem, residue management, soil microbial biomass

Sheibani, S., Yanni, S. F., Wilhelm, R., Whalen, J. K., Whyte, L. G., Greer, C. W. et Madramootoo, C. A. 2013. **Bactéries et archaebactéries du sol dans les écosystèmes agricoles associés à la culture prolongée du maïs (*Zea mays* L.) au Québec (Canada).** *Can. J. Soil Sci.* **93**: 45–57. La microflore du sol contrôle les processus biologiques dans le sol et on estime qu'elle donne une bonne idée de la santé générale du sol. Il est particulièrement important d'évaluer la microflore des sols exploités de manière intensive, mais peu travaillés, car les microorganismes sont ceux qui décomposent le plus les résidus abondants qu'on trouve dans les systèmes de ce genre. Les auteurs ont examiné la biomasse microbienne ainsi que la diversité des bactéries et des archaebactéries dans un écosystème agricole constitué d'un gleysol dystrique de type loam sablonneux utilisé pour la culture à long terme (15 ans) du maïs (*Zea mays* L.), au Québec (Canada). Le sol avait été travaillé, travaillé de façon réduite ou non travaillé pour deux apports de résidus (élevé : +R et faible : –R). Pour les analyses, on a recouru au dosage du C et du N de la biomasse microbienne (MBC, MBN), à la numération des cellules par dépôt de rapporteur catalysé-hybridation *in situ* en fluorescence (CARD-FISH) et avec le 5-(4, 6-dichlorotriazinyl) amino fluorescein hydrochloride (DTAF), à l'application de la réaction en chaîne de la polymérase-électrophorèse en gradient de dénaturation (PCR-DGGE) à l'ARNr 16S et à une bibliothèque de clones d'archaebactéries. L'analyse par PCR-DGGE a permis d'établir que les protéobactéries, les actinobactéries et les firmicutes sont les groupes dominants dans tous les cas où il y a travail du sol et gestion des résidus. Les archaebactéries sont diversifiées, la plupart des spécimens identifiés appartenant au phylum Crenarchaeota. Les auteurs ont également dépisté des archaebactéries telluriques appartenant au nouveau phylum Thaumarchaeota, les archaéota chimiolithoautotrophes qui oxydent l'ammoniaque, dans un écosystème agricole du maïs au Québec (Canada). Le traitement +R augmente la biomasse microbienne, si l'on se fie à la

**Abbreviations:** CARD-FISH, catalyzed reporter deposition-fluorescence in situ hybridization; CT, conventional tillage; DTAF, 5-(4, 6-dichlorotriazinyl) amino fluorescein hydrochloride; MBC, microbial biomass carbon; MBN, microbial biomass nitrogen; NT, no tillage; PCR-DGGE, polymerase chain reaction-denaturing gradient gel electrophoresis; –R, residue removed; +R, residue retained; RT, reduced tillage; TOC, total organic carbon

<sup>4</sup>Corresponding author (e-mail: joann.whelen@mcgill.ca).

concentration de BMC et à la numération directe des cellules. Compte tenu des résultats obtenus avec la technique CARD-FISH (numération de bactéries et d'archaebactéries sans les cellules fongiques) et de la teneur en BMC (biomasse microbienne totale, avec les cellules fongiques), les auteurs pensent que la biomasse fongique est plus importante dans les parcelles au sol non travaillé.

**Mots clés:** Phylogénie des archaebactéries, travail de conservation du sol, écosystème agricole associé au maïs, gestion des résidus, biomasse microbienne du sol

The soil microbial community controls all the biological processes that define soil health and quality, soil organic matter content and quality, nutrient availability and cycling, greenhouse gas emissions, and stabilization of carbon (C) in the soil. Soil microorganisms can be useful indicators of soil health and sustainability (Doran and Zeiss 2000) and can be assessed by parameters such as microbial abundance and diversity in a given soil. Being the primary decomposers of organic matter and plant litter, their function is expected to be more important in low impact management systems that promote soil sustainability such as reduced tillage and increased residue input. In such cases, the decomposition function of the microbial community largely determines the availability of nutrients for the crop. The assessment of the microbial community in intensively cropped soils is pertinent, especially with the increase in the adoption of reduced tillage management systems in recent years. Corn (*Zea mays* L.) for grain and fodder, produced on about 416 000 ha in Quebec, Canada (Statistics Canada 2011a, b), constitutes about 20% of crop land in the province. Despite this, there is a paucity of information on the microbial community in corn fields, particularly of the archaeal community that likely contributes to ammonia oxidation in the soil nitrogen (N) cycle (Martens-Habbena et al. 2009).

Tillage systems have a direct effect on soil structure and soil organic matter dynamics (Vyn and Raimbault 1993; Paustian et al. 1997), which are expected to impact the soil microbial biomass, community structure and diversity. Low-impact systems include reduced tillage (RT) that is cultivated by a harrow to a shallower depth and less frequently than the conventional system (CT), and the no-till (NT) system that is minimally disturbed with the residue left intact on the soil surface. These tillage systems provide a good opportunity to investigate the soil microbial community because of their contrasting habitats that might support certain types of organisms more than others. Microbial biomass and diversity are reported to be greater in RT and NT systems compared with CT (Granatstein et al. 1987; Lupwayi et al. 1998; Govaerts et al. 2007; Van Groenigen et al. 2010; Lupwayi et al. 2012) as a result of less disturbance and more residue left after harvesting. Moreover, the reported differences between these systems are usually detected in the top 0–5 cm layer of soil (Lupwayi et al. 2004) where the difference in residue availability between CT and conservation tillage (i.e., RT and NT systems) is most pronounced.

The objective of this study was to investigate the abundance of soil microorganisms and identify the dominant bacteria and archaea in a corn agroecosystem with long-term tillage and residue management treatments in an intensively cropped corn field in Quebec. The effects of tillage and residue management on the microbial abundance were also assessed. The dominant bacteria and archaea were detected with 16S rRNA PCR-DGGE and an archaeal clone library while microbial abundance was assessed with CARD-FISH, direct microscopic cell counts, and microbial biomass by chloroform fumigation–direct extraction.

## MATERIALS AND METHODS

### Site Description

The experimental site was located at the Macdonald Research Farm of McGill University in Ste-Anne de Bellevue, Quebec, Canada (lat. 45°30'N, long. 73°35'W, elevation 35.7 m). Mean annual temperature is 6.2°C with 979 mm of precipitation, based on climate data from the nearby Pierre Elliott Trudeau Airport in Dorval, Quebec, from 1971 to 2000 (Environment Canada 2010). The soil was a Dystric Gleysol of the St-Amable and Courval series with a sandy-loam texture (815 g kg<sup>-1</sup> sand, 89 g kg<sup>-1</sup> silt, and 96 g kg<sup>-1</sup> clay) and 19.9 g organic-C kg<sup>-1</sup> in the top 20 cm. Additional site description was provided by Burgess et al. (1996) and Dam et al. (2005). The site where the soil was sampled was designed as a factorial experiment with three types of tillage (NT, RT and CT) and two levels of residue input [low: –R, <15 cm of corn stalk plus roots; and high: +R, received corn stover (leaves, stalks) and roots, which accounted for an extra 8.2 to 9.8 Mg ha<sup>-1</sup> of residue (dry matter basis) in 2006 and 11.3 to 12.2 Mg ha<sup>-1</sup> of residue (dry matter basis) in 2007; Halpern et al. 2010]. There were three replicates of each treatment, giving 18 experimental plots in total. The site was plowed in the spring of 1991 when the agronomic study was established and planted with continuous corn; details of all cultivation, fertilization and herbicide applications is detailed in Burgess et al. (1996), Dam et al. (2005) and Halpern et al. (2010).

### Community DNA Extraction and Amplification of 16S rRNA Genes and DGGE Procedures

The composition of the soil bacterial and archaeal communities was investigated with 16S rRNA gene-targeted PCR-DGGE on soil samples (0–5 cm depth) collected in October 2006, May 2007, August 2007 and



October 2007. Soils were kept at  $-20^{\circ}\text{C}$  until analysis. The PCR primers, mixtures (also see Whyte and Greer 2005) and operating conditions are summarized in Table 1. Soil DNA was extracted from 1 g of field-moist, frozen soil using the UltraClean™ Soil DNA Kit (Mobio Laboratories, Solana Beach, CA) according to manufacturer instructions. The extracted DNA was diluted 10-fold, and 5  $\mu\text{L}$  of 1:10 DNA dilution containing 50–90 ng DNA (quantified with a NanoDrop ND-1000 spectrophotometer, NanoDrop Technologies, Wilmington, DE) was used as a template DNA for bacterial and archaeal 16S rRNA genes amplification. Primers used in this study were purchased from MWG-Biotech (Huntsville, AL) and prepared as 10  $\mu\text{M}$  solutions in PCR water. All PCR reactions were achieved in 50  $\mu\text{L}$  volumes by adding PCR water to the PCR mixture.

For DGGE analysis, 800 ng of bacterial and archaeal amplicons were applied to an 8% (wt/vol) acrylamide gel containing a 35 to 65% (bacteria and archaea) or 40 to 60% (bacterial group specific) denaturing gradient, as described by Perreault et al. (2007). All DGGE analyses were performed in two replicates to ensure reproducibility of results. Comparisons of DGGE banding patterns were made with GelCompar software ([www.applied-maths.com](http://www.applied-maths.com)) by constructing dendrograms with the UPGMA method for grouping and the Jaccard coefficient of similarity.

Selected DGGE bands (120 in total, 67 sent for sequencing) were excised from the bacterial and archaeal gels with a cutting tip (DiaMed Laboratory, Mississauga, ON) and eluted in 50  $\mu\text{L}$  of water at  $4^{\circ}\text{C}$  overnight and re-amplified with the appropriate corresponding bacteria and archaea primers (Table 1) without the guanine-cytosine (GC) clamps to optimize the PCR reaction. The PCR products from bacterial DGGE were purified with a Qiagen gel purification kit (Mississauga, ON) to remove salts, dNTPs, primers, and enzymes following the manufacturer's protocol, cloned with the pGEM-T Easy vector system (Promega, Madison, WI) following Steven et al. (2007) and sent for sequencing at the Laval University Bioinformatics Centre. No bands were excised from  $\alpha$ -Proteobacteria and  $\beta$ -Proteobacteria DGGE gels for sequencing. The archaeal 16S rRNA fragments were sequenced directly without cloning at the Laval University Bioinformatics Centre.

The 16S rRNA gene fragments in isolated DGGE bands and clones were identified by comparing to known DNA sequences in the NCBI, GenBank database using the BLASTn algorithm (Altschul et al. 1990). Taxonomic affiliations of sequences were also determined using the RDP classifier function of the Ribosomal Database Project-II release 9 with a confidence threshold of 80% (<http://rdp.cme.msu.edu/classifier>). Nucleotide sequences from this study were submitted to the GenBank database under accession numbers JF303758–JF303854.

### Phylogenetic Analysis of Archaeal 16S rRNA Clone Library

An archaeal clone library was constructed to help identify individual populations in the community, which was expected to complement and confirm the identification obtained from DGGE sequences. An RT+R soil sample collected during the growing season (August 2007) was selected because the DGGE profile showed that this sample harbored high archaeal diversity, relative to all other samples. The library was constructed as described by Steven et al. (2007), using the 751F and 1406R primers (Patricia et al. 2006). For restriction fragment length polymorphism (RFLP), a double digestion of PCR amplified clone DNA was performed using the 4-mer restriction endonucleases *RsaI* and *HhaI* (Invitrogen, Burlington, ON). Digestion reactions were composed of 7.6  $\mu\text{L}$  of PCR water, 0.3  $\mu\text{L}$  of each of 10 U  $\mu\text{L}^{-1}$  *RsaI* and *HhaI*, 1.8  $\mu\text{L}$  of 10  $\times$  REACT I buffer (50 mM Tris-HCl (pH 8.0) and 10 mM  $\text{MgCl}_2$ ) (Invitrogen, Carlsbad, CA), and 8  $\mu\text{L}$  of PCR product. Reactions were incubated at  $37^{\circ}\text{C}$  for 3 h. To inactivate the enzymes, reactions incubated further at  $65^{\circ}\text{C}$  for 20 min. Electrophoresis was carried out using a Sub-Cell Model 96 Agarose Gel Electrophoresis System (Bio-Rad Laboratories, Hercules, CA) for 3 h at 60 V in 1  $\times$  TAE buffer. Gels were stained by adding 1  $\mu\text{L}$  of ethidium bromide (EtdBr) to the gel directly. Gels were viewed under UV light (Bio-Rad Laboratories, Hercules, CA) and images were captured with the associated GeneSnap software (SynGene, Fredrick, MD). A Neighbor-Joining tree was constructed from a ClustalW alignment (across 251 bp of shared sequence) using MEGA software (version 5.05) (Tamura et al. 2011), included bootstrap analysis with 10 000 replications. Rarefaction analysis, library coverage and diversity indices were performed using the DOTUR program (computer algorithms for defining operational taxonomic units and estimating species richness) at 95% sequence similarity (Schloss and Handelsman 2005).

### Microbial Biomass and Direct Counts of Bacteria and Archaea

Composite soil samples (about 300 g subsamples, from five randomly selected locations within the plot) were taken, from the 0–5 cm soil layer, after corn harvest, before fall tillage operations in October 2006 and October 2007 for microbial biomass by chloroform fumigation–direct extraction (kept at  $+4^{\circ}\text{C}$  and processed within 2–7 d; we assumed that microbial cell growth during storage was uniform across treatments). Samples for CARD-FISH and direct cell counts were taken in October 2006 and May, August, and October 2007 and transferred immediately to a freezer at  $-20^{\circ}\text{C}$  until analysis.

The chloroform fumigation–direct extraction method (Voroney et al. 2008) was used to assess microbial biomass C and N concentrations. Paired subsamples were either directly extracted with 0.5 M  $\text{K}_2\text{SO}_4$



Table 1. Description of PCR primers and condition used in this study

No.	PCR type	Primers	PCR mixture	PCR condition
1	Bacteria 16S rRNA: regular PCR	Universal primers: 341F+GC/758R (Steven et al. 2008)	5 $\mu$ L PCR 10 $\times$ buffer with 1.5 mM MgCl <sub>2</sub> , (Qiagen Canada), 0.5 $\mu$ L BSA 10%, 1 $\mu$ L of 10 mM dNTPs mixture, 2.5 $\mu$ L of each 10 $\mu$ M primer (341F+GC/758R), 0.2 $\mu$ L of HotStar <i>Taq</i> DNA Polymerase (5U $\mu$ L <sup>-1</sup> ) (Qiagen, Canada) and 5 $\mu$ L of template DNA	Denaturation: 95°C (15 min), 10 touchdown cycles (65–55°C), 25 cycles at 55°C (30 s), 72°C (1 min and 30 s), final extension: 72°C (10 min)
2	Bacteria 16S rRNA: nested PCR	$\alpha$ - Proteobacteria primers: Alf28F/Alf684R (Muhling et al. 2008)	5 $\mu$ L PCR 10 $\times$ buffer with 1.5 mM MgCl <sub>2</sub> , 0.5 $\mu$ L BSA 10%, 1 $\mu$ L of 10 mM dNTPs mix, 1 $\mu$ L of each 10 $\mu$ M primer (Alf28F/Alf684R), 0.2 $\mu$ L of HotStar <i>Taq</i> DNA Polymerase (5U $\mu$ L <sup>-1</sup> ) and 5 $\mu$ L of template DNA	First PCR: 95°C (15 min), 35 cycles at 95°C (1 min), annealing temperature of 65°C (1 min), 72°C (1 min), final extension: 72°C (10 min) Second PCR: 5 $\mu$ L of group specific PCR products (dilution 1:100) were used as template in a re-PCR with 341F+GC/518R primers, under same PCR conditions (except annealing temperature was 55°C)
3	Bacteria 16S rRNA: semi-nested PCR	$\beta$ -Proteobacteria primers: Beta359F/Beta682R (Muhling et al. 2008)	5 $\mu$ L PCR 10 $\times$ buffer with 1.5 mM MgCl <sub>2</sub> , 0.5 $\mu$ L BSA 10%, 1 $\mu$ L of 10 mM dNTPs mix, 1 $\mu$ L of each 10 $\mu$ M primer (Beta359F/Beta682R), 0.2 $\mu$ L of HotStar <i>Taq</i> DNA Polymerase (5U $\mu$ L <sup>-1</sup> ), and 5 $\mu$ L of template DNA	First PCR: 95°C (15 min), 35 cycles at 95°C (1 min), annealing temperature of 60°C (1 min), 72°C (1 min), final extension: 72°C (10 min) Second PCR: 5 $\mu$ L of group specific PCR products (dilution 1:100) were reamplified using the primer pair 518F+GC/Beta682R, under same PCR conditions (except annealing temperature was 57°C)
4	Archaea 16S rRNA: Semi-nested PCR	First PCR: 109F (Bottos et al. 2008)/934R (Perreault et al. 2007) Second PCR: 344F+GC/934R (Bottos et al. 2008)	First and second PCR as describe for No.1 except for the 0.5 $\mu$ L of HotStar <i>Taq</i> DNA Polymerase (5U $\mu$ L <sup>-1</sup> )	First PCR: denaturation: 95°C (15 min), 35 cycles at 95°C (1 min), annealing: 60°C (1 min), extension: 72°C (1 min), final extension: 72°C (10 min) Second PCR: as the first PCR except for annealing temperature at 57°C (1 min)



**Fig. 1.** A representative denaturing gradient gel electrophoresis (DGGE) analysis of Bacterial 16S rRNA genes amplified from the following samples: RT+R (0–5 cm): A, after harvest (Nov. 2006); B, in spring (May 2007); C, during growing season (Aug. 2007); D, after harvest (Oct. 2007). RT+R (5–20 cm): E, Nov. 2006; F, May 2007; G, Aug. 2007; H, Oct. 2007. NT–R (5–20 cm): I, Nov. 2006; J, May 2007; K, Aug. 2007; L, Oct. 2007. DGGE was run on an 8% acrylamide gel with a gradient of urea and formamide from 35 to 65%.

(1:4 soil:extractant) or kept at 20°C and fumigated with chloroform for 5 d before extraction. Analytical blanks (empty beakers without soil) were used to correct the baseline C and N concentrations. Microbial biomass C was the difference in total organic carbon (TOC, determined using a Shimadzu TOC-V analyzer, Shimadzu Corporation, Kyoto, Japan) concentration of fumigated and un-fumigated extracts divided by an efficiency factor  $k_{EC}=0.45$  (Joergensen 1996). For MBN, the fumigated and un-fumigated extracts underwent persulfate digestion (Cabrera and Beare 1993) and colorimetric analysis for  $\text{NH}_4\text{-N}$  and  $\text{NO}_3\text{-N}$  using a Lachat Quik Chem AE flow injection autoanalyzer (Lachat Instruments, Milwaukee, WI). Microbial biomass N was the difference in mineral N concentration between fumigated and un-fumigated extracts, divided by the efficiency factor  $k_{EN}=0.54$  (Joergensen and Mueller 1996).

The CARD-FISH was performed as described by Niederberger et al. (2010) using the probes EUB338 and ARCH915 (Dijk 2008) with the following modifications:

we used 0.25 g (dry weight basis) of frozen, field-moist soil. Fixed samples were diluted a 1000 times using sodium phosphate buffer, dispersed by ultrasound at  $300 \text{ J mL}^{-1}$  for 20 s and then filtered. No proteinase K digestion was performed. The filter was viewed using a Nikon Eclipse E600 microscope at  $1000\times$  magnification. Bacterial or archaeal cells, stained green against the red background, were counted in a minimum of 10 fields of view for each sample and their numbers calculated following the procedure of Bhupathiraju et al. (1999).

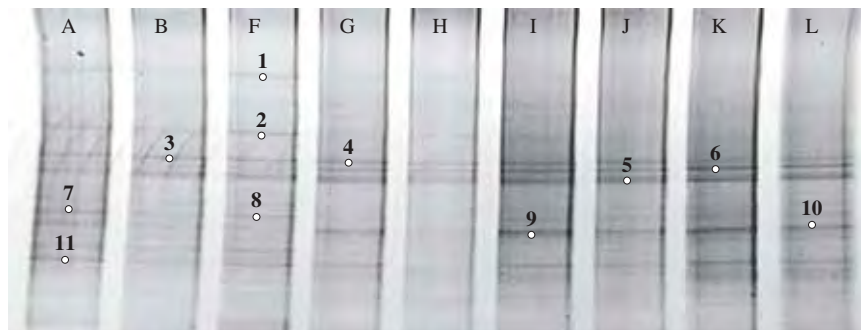
Direct microbial counts were also measured by fluorescence microscopy after staining with 5-(4, 6-dichlorotriazinyl) amino fluorescein hydrochloride (DTAF) following the procedure of Steven et al. (2007). Soil fixation, sonication and bacterial enumeration were performed as mentioned above for CARD-FISH analysis.

### Statistical Analysis

As the field experiment was arranged in a completely randomized design with one replicate per block, the



**Fig. 2.** Dendrogram of denaturing gradient gel electrophoresis (DGGE) of Bacterial 16S rRNA communities for the samples collected during the growing season (Aug. 2007). A gradient runs from left to right at 40–60% on 8% acrylamide for Bacterial gel. The scale of the dendrogram was given as percent of similarity.



**Fig. 3.** A representative denaturing gradient gel electrophoresis (DGGE) analysis of Archaeal 16S rRNA genes amplified from samples collected during the growing season (Aug. 2007). A: RT+R (0–5 cm); B: RT+R (5–20 cm); F: NT+R (5–20 cm); G: CT+R (0–5 cm); H: CT+R (5–20 cm); I: RT–R (0–5 cm); J: RT–R (5–20 cm); K: CT–R (0–5 cm); L: CT–R (5–20 cm). DGGE was run on an 8% acrylamide gel with a gradient of urea and formamide from 35 to 65%.

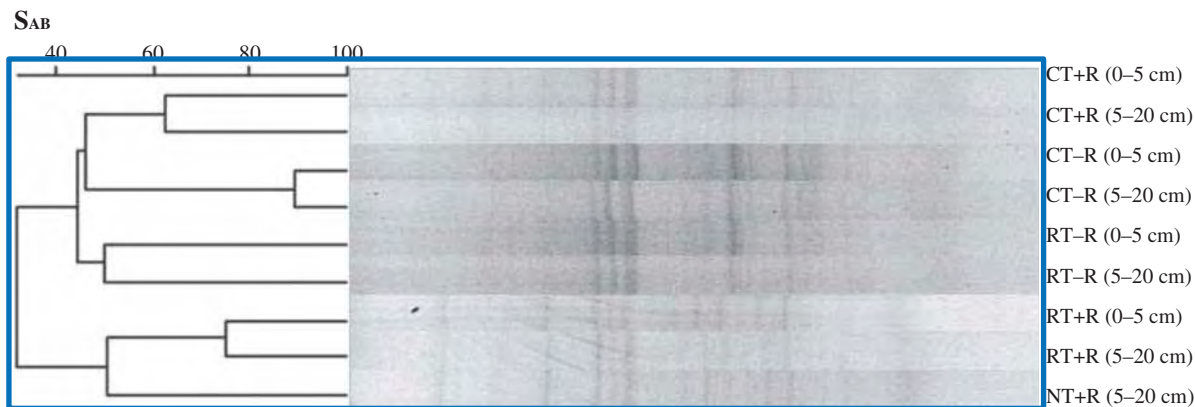
main effects of the residue, tillage and block and the tillage by residue interaction, on microbial biomass C and N, CARD-FISH and direct cell counts were tested using the Proc GLM procedure on SAS statistical software (SAS Institute, Inc. 2009). All data were tested for normality prior to analysis of variance. Least square means with the Tukey adjustment were calculated and significance reported at the 95% confidence level ( $P < 0.05$ ).

## RESULTS AND DISCUSSION

### Bacteria and Archaeal Analysis by PCR-DGGE

Bacterial universal primers, 341F-GC and 758R (Table 1), amplified a ~400 bp region of the Bacterial 16S rRNA gene, which included a ~40 bp GC clamp. Visual observation of DGGE profiles showed many similarities in banding pattern among all of the analyzed samples from two different depths (Fig. 1). This pattern was repeated over the four sampling dates with no noticeable changes. The banding patterns indicated that the structure of the bacterial community at different depths was fairly constant and it did not change significantly as a consequence of tillage and residue

treatments. Most of the observed bands seemed to be shared between samples and no major bands were observed to appear or disappear (Fig. 1). All DGGE profiles were characterized by the presence of two strong bands residing in the low gradient region and a larger number of weaker bands (Fig. 1). Dendrogram analysis was used to compare the banding patterns. The bacterial community differentiated either into two or three clusters (Fig. 2); cluster analyses for the samples collected after harvest (November 2006) and the following spring (May 2007) showed that DGGE banding patterns divided into two clusters based on their depth (0–5 cm and 5–20 cm). The similarity index ( $S_{AB}$ ) value between the two depths of sampling was low ( $S_{AB}$ : 34% and 19.7% respectively), indicating distinctly different communities (Fig. 2). For samples collected during the growing season (August 2007) and after harvest in October 2007, three groups were observed with low similarity index ( $S_{AB}$ : 17% and 14%, respectively, Fig. 2). Group-specific primers ( $\alpha$ - and  $\beta$ -Proteobacteria) provided a similar banding pattern, but fewer bands, than the bacterial universal primers (data not shown).



**Fig. 4.** Dendrogram of denaturing gradient gel electrophoresis (DGGE) of Archaeal 16S rRNA communities. A gradient runs from left to right at 40–60% on 8% acrylamide for Archaeal gel. The scale of the dendrogram was given as percent of similarity.

**Table 2. Microbial biomass C and N concentrations, CARD-FISH bacterial and archaeal cell counts, and DTAF total counts, in the 0–5 cm depth of plots under continuous corn production with tillage and residue input treatments. Values are the mean  $\pm$  standard error ( $n = 6$  for tillage effect and  $n = 9$  for residue effect)**

Treatment	MBC (mg C kg <sup>-1</sup> )	MBN (mg N kg <sup>-1</sup> )	MBC (mg C kg <sup>-1</sup> )	MBN (mg N kg <sup>-1</sup> )	Bacterial CARD-FISH (cells g <sup>-1</sup> soil)	Archaeal CARD-FISH (cells g <sup>-1</sup> soil)	DTAF total cell counts (cells g <sup>-1</sup> soil)
	2006		2007				
NT	342 $\pm$ 43.2a	58 $\pm$ 9.7	351 $\pm$ 36.7a	63.4 $\pm$ 22.6	7.50 $\times$ 10 <sup>8</sup> $\pm$ 3.74 $\times$ 10 <sup>7</sup> a	7.97 $\times$ 10 <sup>7</sup> $\pm$ 2.94 $\times$ 10 <sup>6</sup> a	3.08 $\times$ 10 <sup>8</sup> $\pm$ 1.65 $\times$ 10 <sup>7</sup> a
RT	244 $\pm$ 66.1ab	32 $\pm$ 10	290 $\pm$ 47.3ab	43.3 $\pm$ 10.2	9.15 $\times$ 10 <sup>8</sup> $\pm$ 4.02 $\times$ 10 <sup>7</sup> b	8.61 $\times$ 10 <sup>7</sup> $\pm$ 4.78 $\times$ 10 <sup>6</sup> a	4.26 $\times$ 10 <sup>8</sup> $\pm$ 1.93 $\times$ 10 <sup>7</sup> b
CT	130 $\pm$ 5.63b	36 $\pm$ 3.0	201 $\pm$ 16.3b	20.4 $\pm$ 2.80	7.06 $\times$ 10 <sup>8</sup> $\pm$ 5.14 $\times$ 10 <sup>7</sup> a	6.23 $\times$ 10 <sup>7</sup> $\pm$ 5.77 $\times$ 10 <sup>6</sup> b	3.20 $\times$ 10 <sup>8</sup> $\pm$ 4.23 $\times$ 10 <sup>7</sup> a
-R	175 $\pm$ 31.1a	34 $\pm$ 7.3	232 $\pm$ 27.2a	24.2 $\pm$ 2.95	7.51 $\times$ 10 <sup>8</sup> $\pm$ 5.34 $\times$ 10 <sup>7</sup>	7.29 $\times$ 10 <sup>7</sup> $\pm$ 6.19 $\times$ 10 <sup>6</sup>	3.06 $\times$ 10 <sup>8</sup> $\pm$ 2.67 $\times$ 10 <sup>7</sup> a
+R	302 $\pm$ 65.2b	50 $\pm$ 10	330 $\pm$ 34.9b	60.5 $\pm$ 15.7	8.29 $\times$ 10 <sup>8</sup> $\pm$ 3.38 $\times$ 10 <sup>7</sup>	7.92 $\times$ 10 <sup>7</sup> $\pm$ 3.39 $\times$ 10 <sup>6</sup>	3.97 $\times$ 10 <sup>8</sup> $\pm$ 2.19 $\times$ 10 <sup>7</sup> b
	<i>Treatment effects</i>						
Tillage	$P = 0.0042^{**}$	NS	$P = 0.0080^{**}$	NS	$P = 0.0063^{**}$	$P = 0.0074^{**}$	$P = 0.0005^{**}$
Residue	$P = 0.0085^{**}$	NS	$P = 0.0096^{**}$	NS	NS	NS	$P = 0.0040^{**}$
Tillage $\times$ Residue	NS	NS	NS	NS	NS	NS	$P = 0.0114^{*}$

a, b Within a column, values with the same letter within a treatment (tillage, residue input) did not differ significantly at  $P < 0.05$  (Tukey test).  
\*, \*\* Treatment effects significant at  $P < 0.05$  and  $P < 0.01$ , respectively; NS, non-significant.

The PCR-DGGE analysis of Archaeal 16S rRNA PCR-DGGE revealed 6 to 15 bands of relatively uniform intensity and with similar banding patterns (Fig. 3). There were more bands during the growing season (August 2007) than in spring (May 2007) or after harvest (November 2006 and October 2007), but cluster analysis did not reveal clear trends with respect to 16S rRNA sequences in the various treatments or at different sampling dates (Fig. 4).

### Bacteria and Archaeal Diversity

Visualization limitations prevented the physical isolation of all bands in the DGGE gels, but bacterial 16S rRNA gene sequences of 38 bands were compared with known sequences in GenBank using the BLAST algorithm. This analysis showed that bands belonged to the phyla Proteobacteria (25 bands), Actinobacteria (7 bands), Firmicutes (3 bands) and Acidobacteria (2 bands), which are common in agricultural soils (Table 3). Proteobacteria and actinobacteria are usually reported to be amongst the most abundant phyla in soils (e.g., Amann et al. 1995; Buckley and Schmidt 2001; Valinsky et al. 2002; Upchurch et al. 2008). Proteobacteria are a major phylum in soil, with high morphological, physiological and metabolic diversity that permits them to participate in global carbon, nitrogen and sulfur cycling; this group includes N<sub>2</sub> fixing and ammonia oxidizing bacteria (Kersters et al. 2006). Soil actinobacteria, mostly aerobic, are known to be important decomposers, particularly of cellulose, the most abundant material in plant litter (Upchurch et al. 2008). Some members of the Actinobacteria form branching filaments similar to fungal mycelia, which in theory may make them susceptible to disturbance by tillage, similar to fungi (Upchurch et al. 2008). The Firmicutes are also abundant in soils, with aerobic and anaerobic free-living or pathogenic members; some are decomposers of organic matter and many are resilient because of their ability to survive extreme dry conditions by endospore formation (Cleveland et al. 2007; Ludwig et al. 2009). The functions of the newly established phylum, the Acidobacteria, are not yet fully clear but they are diverse, ubiquitous and abundant in soils and therefore are thought to have important ecological functions (Barns et al. 1999; Quaiser et al. 2003).

There were no major differences between the profiles with respect to tillage and residue management treatments at the four sampling dates in this study, which is consistent with some reports that showed no effect of tillage and residue management on microbial diversity (e.g., Buckley and Schmidt 2001; Helgason et al. 2010) but is in contrast with others such as Peixoto et al. (2006), Garbeva et al. (2006) and Upchurch et al. (2008), which showed bacterial diversity to be affected by tillage and residue inputs. Since diverse bacterial phyla were detected across tillage and residue treatments, it suggests that bacteria were either minimally affected by tillage and residue inputs, or had the ability to recover quickly



**Table 3. 16S rRNA gene analyses of isolated bands from bacteria DGGE. Sequences were compared with known sequences in GenBank using the BLAST algorithm. Phylogenetic classification of sequences was determined using the RDP classifier function of the Ribosomal Database Project-II release 9**

Band	Closest BLAST match	Origin of BLAST match	Similarity to BLAST match	RDP grouping
1a	Bacterium KMS200711_068 (EU881327)	Maize cropland soil	96% (358/372)	Polyangiaceae 100% (family)
1b	Soil bacterium 4M1-E07 (EU052014)	Savanna soil	96% (391/408)	Bacteria 100% (domain)
2	Bacterium FFCH8343 (EU133494)	Undisturbed soil	99% (406/411)	Rhizobiales 99% (order)
3	Soil bacterium 2_G8 (EU589321)	Rice paddy field soil	92% (399/431)	Betaproteobacteria 100% (phylum)
4	Firmicutes Raunefjorden 04 (AM706659)	Environmental sample	97% (398/411)	Lactobacillales 96% (order)
5a	Bacterium FFCH15545 (EU132916)	Undisturbed soil	92% (372/405)	Actinomycetales 99% (order)
5b	Firmicutes Raunefjorden11 (AM706663)	Environmental samples	99% (401/400)	Lactobacillales 95% (order)
6	Idobacteria GASP-WB1W1_E08 (EF073599)	pasture	98% (386/390)	Acidobacteriaceae 95% (family)
7a	Firmicutes Raunefjorden 11 (AM706663)	Environmental sample	95% (371/390)	Lactobacillales 97% (order)
7b	Soil bacterium clone 2_G9 (EU589322)	Rice paddy field soil	94% (384/408)	Proteobacteria 96% (phylum)
8	Actinomycetales TL1226 (EU699684)	Soil	99% (405/410)	Actinomycetales 100% (order)
9	Micromonospora sp. HBUM49436 (EU119220)	Soil sample	94% (374/397)	Actinomycetales 99% (order)
10	Betaproteobacterium GASP-MA2S2_H05 (EF662904)	Cropland	89% (315/355)	Lactobacillales 97% (phylum)
11	Ramlibacter sp. P-8 (AM411936)	Rice paddy soil	97% (393/404)	Comamonadaceae 96% (family)
12	Ramlibacter tataouinensis 153-3 (AJ871240)	Biological soil crusts	96% (388/401)	Comamonadaceae 92% (family)
13	Alphaproteobacterium OS-C02 (EF612398)	Soil	99% (385/381)	Sphingomonadaceae 98% (family)
14	Agricultural soil bacterium isolate SI-15 (AJ252582)	Agricultural soil bacterium	98% (401/408)	Rhodobacteriaceae 98% (family)
15	Alphaproteobacterium 5kpl2aC11 (EF092525)	Environmental sample	99% (398/394)	Rhodobacteriaceae 98% (family)
16	Deltaproteobacterium g65 (EU979074)	Rhizosphere of faba bean	92% (375/408)	Polyangiaceae 100% (family)
17	Pseudomonas sp.G-229-23 (EF102852)	Rhizosphere of tobacco	98% (402/411)	Pseudomonas 83% (genus)
18	Prophyrobacter sp. AUVE_14G05 (EF651683)	Cropland	94% (361/382)	Sphingomonadaceae 98% (family)
19	Bacterium 4PS16S. (AY365088)	Agricultural soil	98% (387/393)	Pseudomonas 91% (genus)
20	Bacterium 7PS16S. (AY365091)	Agricultural soil	97% (398/410)	Pseudomonas 91% (genus)
21	Acidobacteria GASP-WB1W1_E08 (EF073599)	pasture	98% (386/390)	Gp3 99% (genus)
22	Rhizobiales bacterium AhedenP3 (FJ475499)	Forest soil	97% (326/333)	Rhizobiales 100% (order)
23	Actinomycetales bacterium TL1213 (EU699671)	Soil	98% (390/398)	Streptomycetaceae 98% (family)
24	Gammaproteobacterium GASPMA2S3_D1 (EF663052)	Cropland	97% (401/412)	Proteobacteria 96% (phylum)
25	Hyphomicrobiaceae GASP-KC1S3_A04. (EU299238)	Restored grassland	90% (230/255)	Alphaproteobacteria 100% (class)
26	Hyphomicrobium sp. AUVE_04B07 (EF651116)	Cropland	93% (297/319)	Alphaproteobacteria 92% (class)
27	Soil bacterium W4Ba49 (DQ643713)	Agricultural soil	97% (358/366)	Alphaproteobacteria 98% (class)
28	Actinobacterium TH1-94 (AM690885)	Environmental sample	98% (404/413)	Actinomycetales 99% (order)
29	Caulobacteriales Plot29-H11_16S (EU202838)	Agricultural soil	97% (326/333)	Sphingosinicella 100% (genus)
30	Kaistobacter sp. Plot03-H09 (EU276575)	Agricultural soil	97% (373/381)	Sphingosinicella 98% (genus)
31	Bacterium FFCH15545 (EU132916)	Undisturbed soil mixed grass	93% (369/396)	Actinomycetales 100% (order)
32	Streptomyces sp. 33D (EF585406)	Soil	97% (395/407)	Streptomycetaceae 98% (family)
33	Bacterium 7PS (AY365091)	Agricultural soil	98% (415/421)	Pseudomonadaceae 83% (family)
34	Sphingomonas sp.GASPMA1W1_C03 (EF662616)	Cropland	99% (316/319)	Sphingomonadaceae 98% (family)

from disturbance. Tillage breaks up larger soil aggregates (>250 µm), fragments and incorporates surface residue, and eliminates some of the larger soil fauna (Paustian et al. 1997), but there is no evidence that it induces a fundamental change the microstructure at the scale of bacterial habitats (within micropores, 10 µm or less); this needs to be evaluated more closely in future studies. The +R plots received between 8 and 12 Mg ha<sup>-1</sup> yr<sup>-1</sup> of extra corn residues, but it was the same type of residue as in the -R plots, so this leads to the expectation of larger bacterial populations (Metting 1993), but not necessarily a shift in diversity because the food quality was the same across the field. We also acknowledge a bias with PCR-based techniques such as DGGE, in that short subunits of rRNA are amplified according to a competitive enzymatic reaction so that the most abundant microbial populations yield the most amplicons, and organisms that are present at 1% or less of the total community might not even appear on the gels, even if their populations exceed 10<sup>4</sup> to 10<sup>5</sup> cells per

gram of soil (Forney et al. 2004). It is possible that bacterial species sensitive to the tillage and residue treatments were not detected due to limitations of the PCR-DGGE method.

Comparing archaeal 16S rRNA sequences to those in GenBank revealed matches with the phylum Crenarchaeote in isolates from agricultural, rhizosphere and forest soils, which were all classified as Thermoprotei (Table 4). Orders of the class Thermoprotei includes the Thermoproteales and Desulfurococcales, which can be primary producers using oxygen, elemental sulfur, sulfate, thiosulfate, sulfite, and nitrate as electron acceptors, and the Sulfolobales, which oxidize sulfur as an energy substrate and are mainly aerobic thermophilic archaea (Huber et al. 2006; Huber and Stetter 2006).

A total of 59 clones were screened by RFLP and grouped by identical restriction pattern (phylotypes). Analysis of the inserts showed three restriction patterns, but sequencing revealed that 26 clones were related to bacteria, due to poor primer specificity for archaea in



**Table 4.** 16S rRNA gene analyses of the isolated bands from archaea DGGE. Sequences were compared with known sequence in GenBank using the BLAST algorithm. Phylogenetic classification of sequences was determined using the RDP classifier function of the Ribosomal Database Project-II release 9

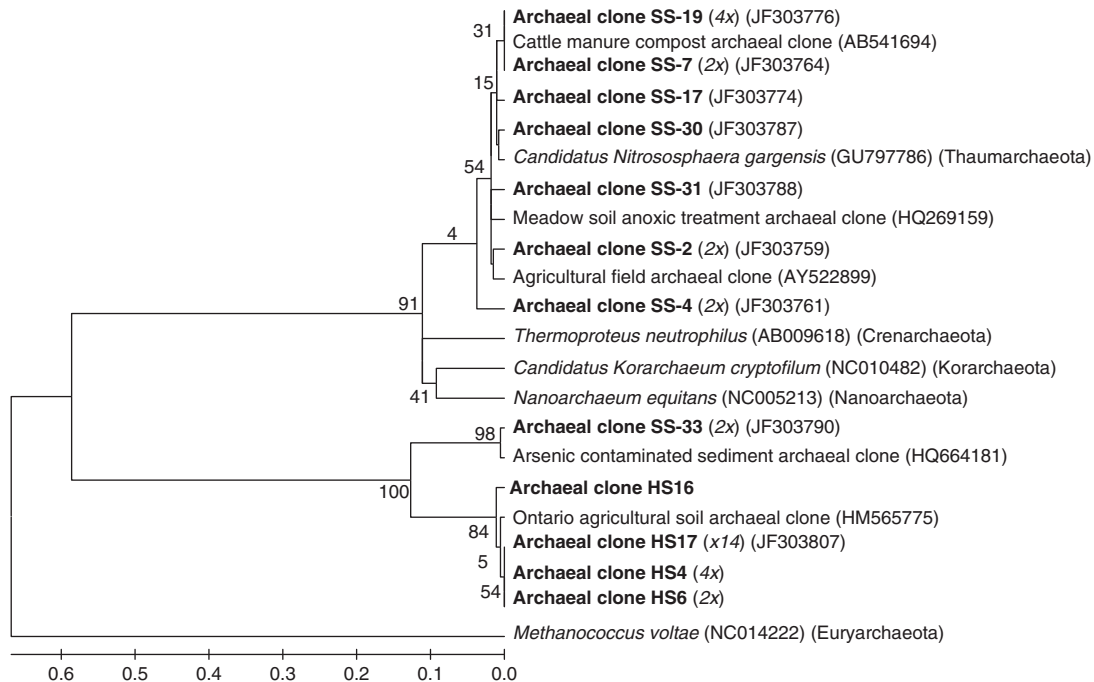
Band	Closest BLAST match	Origin of BLAST match	Similarity to BLAST match	RDP grouping
1	Archaeon clone CAP128RC (EU223281)	Corn rhizosphere soil	98% (441/449)	Thermoprotei 82% (class)
2	Archaeon clone Elev_16S_arch_974 (EF023083)	Rhizosphere	99% (396/400)	Thermoprotei 87% (class)
3	Crenarchaeote clone MBS11 (AY522889)	Mesophilic soil, forest	99% (398/399)	Thermoprotei 81% (class)
4	Crenarchaeote clone MWS36 (AY522861)	Mesophilic soil, turf field	98% (401/409)	Thermoprotei 91% (class)
5	Crenarchaeote clone NRP-M (AB243804)	Rice paddy soil	99% (426/428)	Thermoprotei 85% (class)
6	Crenarchaeote clone A364I-21 (AM292013)	Acidic forest soil	98% (424/432)	Thermoprotei 87% (class)
7	Crenarchaeote clone A364I-08 (AM292000)	Acidic forest soil,	98% (400/407)	Thermoprotei 86% (class)
8	Crenarchaeote clone CBS16S2-2-8 (EF450809)	Agricultural soil	96% (370/384)	Thermoprotei 94% (class)
9	Crenarchaeote clone CBS16S2-2-5 (EF450808)	Agricultural soil	99% (284/286)	Thermoprotei 82% (class)
10	Archaeon clone Elev_16S_arch_999 (EF023106)	Rhizosphere	97% (324/333)	Thermoprotei 85% (class)
11	Crenarchaeote clone MWS36 (AY522861)	Mesophilic soil, turf field	99% (495/499)	Thermoprotei 95% (class)
12	Archaeon clone Elev_16S_arch_945 (EF023055)	Rhizosphere	99% (496/499)	Thermoprotei 96% (class)
13	Crenarchaeote clone A364I-08 (AM292000)	Acidic forest soil	98% (492/498)	Thermoprotei 97% (class)
14	Crenarchaeote clone A364I-08 (AM292000)	Acidic forest soil	97% (481/493)	Thermoprotei 91% (class)
15	Crenarchaeote clone A109-18 (AM291988)	Acidic forest soil	96% (463/479)	Thermoprotei 95% (class)
16	Crenarchaeote clone OdenB-100b (DQ278124)	Soil	98% (481/487)	Thermoprotei 95% (class)
17	Archaeon clone Elev_16_arch_539 (EF022693)	Rhizosphere	99% (493/494)	Thermoprotei 99% (class)
18	Archaeon clone 1-I-12 (EU223277)	Corn rhizosphere soil	98% (489/495)	Thermoprotei 98% (class)
19	Archaeon clone pTN-23 (AB182772)	Rice paddy soil	99% (489/495)	Thermoprotei 96% (class)
20	Archaeon clone pTN-FC-16m (AB182772)	Rice paddy soil	98% (489/495)	Thermoprotei 97% (class)
21	Crenarchaeote clone MBS11 (AY522889)	Mesophilic soil, forest	98% (489/494)	Thermoprotei 92% (class)
22	Crenarchaeote clone CBS16S1-1-2 (EF450802)	Agricultural soil	98% (440/447)	Thermoprotei 81% (class)
23	Archaeon clone Elev_16S_arch_974 (EF023083)	Rhizosphere	98% (487/493)	Thermoprotei 98% (class)
24	Crenarchaeote clone A364I-08 (AM292000)	Acidic forest soil	98% (490/496)	Thermoprotei 93% (class)
25	Crenarchaeote clone MBS11 (AY522889)	Mesophilic soil, forest	99% (494/496)	Thermoprotei 96% (class)
26	Crenarchaeote clone TREC89-24 (AY487106)	Tomato rhizosphere	99% (488/492)	Thermoprotei 98% (class)
27	Crenarchaeote clone MBS11 (AY522889)	Mesophilic forest soil	99% (398/399)	Thermoprotei 81% (class)

agricultural soils. The primer sets selected (Table 1) were developed to detect archaea in the Canadian arctic (Perreault et al. 2007; Bottos et al. 2008) and at the time of this study were the only ones available for Canadian soils. We recommend that researchers select archaea primers developed for temperate agricultural soils to avoid overestimating archaea presence in such soils. Thirty-three clones were reliably identified as archaea from the sequence data, and these were grouped into eight phylotypes. Archaeal clone sequences were assigned to operational taxonomic units at a level of 95% similarity for phylogenetic tree building. Clone coverage with Good's percent coverage, which provides a quantitative estimate of how well the sample size reflects the apparent diversity within the clone library, was 84%. The phylogenetic grouping of the sequences is illustrated in Fig. 1. Archaeal clone library results showed that members of Crenarchaeota were dominant, with one sequence (SS-30) closely related to Thaumarchaeota. Clones classified as Crenarchaeota by RDP did not cluster with any cultured representatives and their top BLASTn matches were with isolates and clones from temperate soil.

The Thaumarchaeota is a newly proposed phylum that was initially classified as Crenarchaeota. In this group are ammonia oxidizing archaea that fix carbon dioxide and oxidize ammonia to nitrite, important functions in the C and N cycles (Leininger et al. 2006;

Spang et al. 2010; Tourna et al. 2011). Leininger et al. (2006) reported that  $\text{NH}_4$ -oxidizing archaea in soils may be more abundant than  $\text{NH}_4$ -oxidizing bacteria. The related  $\text{NH}_4$ -oxidizing marine archaea, *Nitrososphaera maritimus*, and warm water spring archaea *Nitrososphaera gargensis* have a chemolithoautotrophic metabolism, but Tourna et al. (2011) reported that *N. viennensis*, the only  $\text{NH}_4$ -oxidizing archaea isolated from soil to date, showed enhanced growth with the addition of an organic substrate. The detection of close relatives to this  $\text{NH}_4$ -oxidizing archaea in the corn-cropped soil indicates that archaeal groups are likely contributing to nitrification, which produces  $\text{NO}_3^-$  for corn uptake, leaching and denitrification in this soil. Ammonia oxidizers generally catalyze two reactions, converting ammonia to hydroxylamine and then to nitrite, which also produces nitrous oxide ( $\text{N}_2\text{O}$ ) and accounts for its emission in the nitrification process (Whalen and Sampedro 2010). Archaeal ammonia oxides may compete for substrates with bacterial ammonia oxidizers, or perhaps each thrives in differences niches that are suitable to their particular needs. This possibility requires further investigation.

Reports of archaea in upland (oxic) soils are few. Borneman et al. (1996) reported no sequences of archaea in a study on soil microbial diversity in a clover-grass pasture in southern Wisconsin, although they did not clone and sequence their small subunit rDNA



**Fig. 5.** Archaeal 16S rRNA distance based (Neighbor-Joining) phylogenetic tree including bootstrap values for the RT + R (0–5 cm) sample collected during the growing season in Aug. 2007. Clones with greater than 95% sequence similarity were grouped together with the number of each operational taxonomic unit and show the number of clones represented by each branch. Top BLASTn matches are in plain text and cultured representatives from major archaeal clades in italics.

amplification products to confirm the presence or absence of archaea. Yet, Furlong et al. (2002) reported that five patterns constituting a single archaeal lineage were detected in soil and earthworm casts from NT plots at the Horseshoe Bend Long Term Environmental Research site in Georgia, USA. As well, Bintrim et al. (1997) reported a cluster of archaea affiliated with the Crenarchaeota in soils sampled at the West Madison Agricultural Research Station in Wisconsin, USA. Their analysis also showed that the full-length sequences of 10 cloned small subunit ribosomal rRNA genes had phylogenetic affinity to members of the planktonic division of Crenarchaeota.

### Microbial Biomass and Cell Counts

The MBC and MBN concentrations and CARD-FISH and DTAF counts are given in Table 2. The MBC concentration was significantly greater in the NT plots than CT plots and in the +R plots than the –R plots in 2006 and 2007. DTAF cell counts also showed a greater count in the +R plots compared with the –R plots. The MBN trend was similar to MBC but differences were not statistically significant. There was 119% more MBC in the NT plots compared with the CT plots similar to results of Spedding et al. (2004), Granatstein et al. (1987) and Feng et al. (2003), who reported a range between 24 and 140% greater MBC in NT compared with CT. In contrast, the cell counting methods (CARD-FISH and DTAF) showed a somewhat different

trend indicating that RT plots contained more microbial biomass (based on bacterial and archaeal counts) than CT plots. An interaction effect (Table 2) was detected by the DTAF count such that the RT + R plots had greater counts than the NT + R, NT – R, and CT – R plots.

The cell count results suggest that conditions in the RT plots (moderate soil disturbance, some fragmentation and incorporation of crop residues) were more favorable for the development of bacteria and archaea populations than conditions in the CT or NT plots. These methods do not count fungal cells, which are included in MBC and MBN assessments, so cannot be compared directly to the chloroform fumigation–direct extraction method. Taken together, the MBC, CARD-FISH and DTAF results seem to suggest a change in microbial communities leading to greater fungal biomass in NT plots than other tillage systems and larger bacteria and archaea populations in RT than CT plots. Fungal biomass was not directly measured and the fungal to bacterial ratio needs further investigation to verify the abundance of fungal populations in NT systems. Acidic soil conditions are considered to promote fungal populations (Whalen and Sampedro 2010), but soils from NT plots were significantly ( $P < 0.05$ ) less acidic ( $\text{pH} = 6.9 \pm 0.1$ ) than soils from the CT ( $\text{pH} = 6.1 \pm 0.1$ ) and RT ( $\text{pH} = 6.2 \pm 0.1$ ) plots. Our results suggested that the higher MBC concentrations in NT plots were due to more fungal biomass, probably not

related to soil pH conditions but likely related more to the lack of tillage disturbance in NT. We also note that archaea cell counts by CARD-FISH were affected by tillage (NT, RT > CT), which indicates an important presence of archaea in less-disturbed soils plots as well.

The +R plots also exhibited an increase in total bacteria cells with the DTAF method, but there was no difference in the bacteria and archaea populations due to residue management according to the CARD-FISH method. The minimal number of rRNA target molecules required to obtain a visible fluorescence signal after CARD-FISH (with rRNA-targeted probes) is not yet determined and we had difficulty to observe cells with CARD-FISH in soil due to high background fluorescence from the soil itself. This may be a limitation to broader application of the CARD-FISH method to census soil bacteria and archaea populations.

### CONCLUSIONS

This is one of the first reports of archaeal groups in corn agroecosystem soils in Quebec (Canada), as far as we know. One of our sequences was also closely related to Thaumarchaeota, order Nitrososphaerales. Thaumarchaeota has been proposed as a third archaeal phylum. Members of this phyla are ammonia oxidizers and may have different features from those of hyperthermophilic crenarchaeota (Brochier-Armanet et al. 2008). Thaumarchaeota have significant roles in the global C and N cycles and might be the most abundant ammonia oxidizers in soil (Tourna et al. 2011). The extent of the archaeal contribution to microbial functions in this oxic soil is not yet known, but their large population size and known role as ammonia oxidizers (Tourna et al. 2011) in soil suggests that they contribute to the production of plant-available N (e.g., NO<sub>3</sub>) and gaseous emissions (e.g., N<sub>2</sub>O as a byproduct of ammonia oxidation) in our agricultural soils.

Tillage and residue input affected microbial biomass – more microbial biomass in NT plots than CT plots, and greater microbial biomass with higher residue inputs, as indicated by the MBC and supported by the MBN trends. Based on evidence from the chloroform fumigation method (MBC) and the microscopic CARD-FISH and DTAF analysis taken together we propose that tillage affects the microbial community as follows: NT plots are suggested to have a dominant fungal population, with an appreciable archaeal population, RT plots are dominated by bacteria and archaea, and fewer fungi, and the CT plots have the smallest microbial biomass/populations. Soil bacterial diversity was difficult to assess due to few dominant groups (Proteobacteria, Actinobacteria and Firmicutes). Changes in bacterial diversity due to tillage and residue management were not detected by the DGGE method. Microbial community diversity in this study was limited to bacteria and archaeal groups, and it is important also to consider the diversity of fungi, which are more sensitive to tillage disturbance than bacteria (Paustian et al. 1997).

High-throughput DNA sequencing techniques like pyrosequencing prove helpful to assess soil microbial diversity (Liu et al. 2007; Roesch et al. 2007; Jones et al. 2009) and the effect of tillage and crop rotation on the soil microbial community (Yin et al. 2010). Integration of massively parallel pyrosequencing with quantitative real-time PCR (qPCR) holds promise for assessing microbial community structure and functions (Zhang et al. 2011) and could better describe soil microbial responses to long-term agricultural practices like tillage and residue management.

### ACKNOWLEDGEMENTS

We thank Peter Kirby for assistance with the field work. Financial support was from the Fonds de recherche sur la nature et les technologies du Québec.

**Altschul, S., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. 1990.** Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.

**Amann, R. I., Ludwig, W. and Schleifer, K. H. 1995.** Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**: 143–169.

**Barns, S. M., Takala, S. L. and Kuske, C. R. 1999.** Wide distribution and diversity of members of the bacterial kingdom acidobacterium in the environments. *Appl. Environ. Microbiol.* **65**: 1731–1737.

**Bhupathiraju, V. K., Hernandez, M., Krauter, P. and Alvarez-Cohen, L. 1999.** A new direct microscopy based method for evaluating *in-situ* bioremediation. *J. Hazard. Mater.* **67**: 299–312.

**Bintrim, S. B., Donohue, T. J., Handelsman, J., Roberts, G. P. and Goodman, R. M. 1997.** Molecular phylogeny of archaea from soil. *Proc. Natl. Acad. Sci. USA.* **94**: 277–282.

**Borneman, J., Skroch, P., O'Sullivan, K., Palus, J., Rumjanek, N., Jansen, J., Nienhuis, J. and Triplett, E. 1996.** Molecular microbial diversity of an agricultural soil in Wisconsin. *Appl. Environ. Microbiol.* **62**: 1935–1943.

**Bottos, E., Vincent, W., Greer, C. and Whyte, L. 2008.** Prokaryotic diversity of arctic ice shelf microbial mats. *Environ. Microbiol.* **10**: 950–966.

**Brochier-Armanet, C., Boussau, B., Gribaldo, S. and Forterre, P. 2008.** Mesophilic crenarchaeota: Proposal for a third archaeal phylum, the Thaumarchaeota. *Nat. Rev. Microbiol.* **6**: 245–252.

**Buckley, D. and Schmidt, T. 2001.** The structure of microbial communities in soil and the lasting impact of cultivation. *Microb. Ecol.* **42**: 11–21.

**Burgess, M. S., Madramootoo, C. A. and Mehuys, G. R. 1996.** Tillage and crop residue effects on corn production in Quebec. *Agron. J.* **88**: 792–797.

**Cabrera, M. and Beare, M. 1993.** Alkaline persulfate oxidation for determining total nitrogen in microbial biomass extracts. *Soil Sci. Soc. Am. J.* **57**: 1007.

**Cleveland, C. C., Nemergut, D. R., Schmidt, S. K. and Townsend, A. R. 2007.** Increases in soil respiration following labile carbon additions linked to rapid shifts in soil microbial community composition. *Biogeochemistry* **82**: 229–240.

**Dam, R. F., Mehdi, B. B., Burgess, M. S. E., Madramootoo, C. A., Mehuys, G. R. and Callum, I. R. 2005.** Soil bulk density and crop yield under eleven consecutive years of corn with



different tillage and residue practices in a sandy loam soil in Central Canada. *Soil Tillage Res.* **84**: 41–53.

**Dijk, J. A., Breugelmanns, P., Philips, J., Haest, P. J., Smolders, E. and Springael, D. 2008.** Catalyzed reporter deposition-fluorescent *in situ* hybridization (CARD-FISH) detection of dehalococoides. *J. Microbiol. Methods* **73**: 142–147.

**Doran, J. W. and Zeiss, M. R. 2000.** Soil health and sustainability: managing the biotic component of soil quality. *Appl. Soil Ecol.* **15**: 3–11.

**Environment Canada. 2010.** Canadian climate normals 1970–2000: Montreal/Pierre Elliott Trudeau International Airport. Environment Canada, Ottawa, ON. [Online] Available: [http://climate.weatheroffice.gc.ca/climate\\_normals/index\\_e.html](http://climate.weatheroffice.gc.ca/climate_normals/index_e.html) [2010 Oct. 22].

**Feng, Y., Motta, A. C., Reeves, D. W., Burmester, C. H., van Santen, E. and Osborne, J. A. 2003.** Soil microbial communities under conventional-till and no-till continuous cotton systems. *Soil Biol. Biochem.* **35**: 1693–1703.

**Forney, L. J., Zhou, X. and Brown, C. J. 2004.** Molecular microbial ecology: land of the one-eyed king. *Curr. Opin. Microbiol.* **7**: 210–220.

**Furlong, M., Singleton, D., Coleman, D. and Whitman, W. 2002.** Molecular and culture-based analyses of prokaryotic communities from an agricultural soil and the burrows and casts of the earthworm *Lumbricus rubellus*. *Appl. Environ. Microbiol.* **68**: 1265–1279.

**Garbeva, P., Postma, J., Van Veen, J. and Van Elsas, J. 2006.** Effect of above ground plant species on soil microbial community structure and its impact on suppression of *Rhizoctonia solani* AG3. *Environ. Microbiol.* **8**: 233–246.

**Govaerts, B., Mezzalama, M., Unno, Y., Sayre, K., Luna-Guido, M., Vanherck, K., Dendooven, L. and Deckers, J. 2007.** Influence of tillage, residue management, and crop rotation on soil microbial biomass and catabolic diversity. *Appl. Soil Ecol.* **37**: 18–30.

**Granatstein, D. M., Bezdicke, D. F., Cochran, V. L., Elliott, L. F. and Hammel, J. 1987.** Long-term tillage and rotation effects on soil microbial biomass, carbon and nitrogen. *Biol. Fertil. Soils* **5**: 265–270.

**Halpern, M. T., Whalen, J. K. and Madramootoo, C. A. 2010.** Long-term tillage and residue management influences soil carbon and nitrogen dynamics. *Soil Sci. Soc. Am. J.* **74**: 1211–1217.

**Helgason, B. L., Walley, F. L. and Germida, J. J. 2010.** Long-term no-till management affects microbial biomass but not community composition in Canadian prairie agroecosystems. *Soil Biol. Biochem.* **42**: 2192–2202.

**Huber, H., Huber, R. and Stetter, K. 2006.** Thermoproteales. Pages 10–12 in M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, and E. Stackebrandt, eds. *The Prokaryotes*. Springer, New York, NY.

**Huber, H. and Stetter, K. 2006.** Desulfurococcales. Pages 52–68 in M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, and E. Stackebrandt, eds. *The prokaryotes*. 3rd ed. Vol. 5. Springer, New York, NY.

**Joergensen, R. G. 1996.** Quantification of the microbial biomass by determining ninhydrin-reactive N. *Soil Biol. Biochem.* **28**: 301–306.

**Joergensen, R. G. and Mueller, T. 1996.** The fumigation-extraction method to estimate soil microbial biomass: calibration of the  $k_{EN}$  value. *Soil Biol. Biochem.* **28**: 33–37.

**Jones, R. T., Robeson, M. S., Lauber, C. L., Hamady, M., Knight, R. and Fierer, N. 2009.** A comprehensive survey of

soil acidobacterial diversity using pyrosequencing and clone library analyses. *ISME J.* **3**: 442–453.

**Kerstens K., De Vos P., Gillis M., Swings J., Vandamme P. and Stackebrandt E. 2006.** Introduction to the proteobacteria. Pages 3–37 in M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, E. and Stackebrandt, eds. *The prokaryotes*. 3rd ed. Vol. 5. Springer, New York, NY.

**Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J. and Nicol, G. W. et al. 2006.** Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442**: 806–809.

**Liu, Z.-Zh., Lozupone, C., Hamady, M., Bushman, F. D. and Knight, R. 2007.** Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acid Res.* **35**: 1–10.

**Ludwig, W., Schleifer, K.-H. and Whitman, W. B. 2009.** Revised road map to the phylum Firmicutes. Pages 1–13 in P. De Vos, G. Garrity, D. Jones, N. R. Krieg, W. Ludwig, F. A. Rainey, K.-H. Schleifer, and W. B. Whitman, eds. *Bergey's manual of systematic bacteriology*. 2nd ed. Vol. 3. The firmicutes. Springer, New York, NY.

**Lupwayi, N. Z., Lafond, G. P., Ziadi, N. and Grant, C. A. 2012.** Soil microbial response to nitrogen fertilizer and tillage in barley and corn. *Soil Tillage Res.* **118**: 139–146.

**Lupwayi, N. Z., Clayton, G. W., O'Donovan, J. T., Harker, K. N., Turkington, T. K. and Rice, W. A. 2004.** Soil microbiological properties during decomposition of crop residues under conventional and zero tillage. *Can. J. Soil Sci.* **84**: 411–419.

**Lupwayi, N. Z., Rice, W. A. and Clayton, G. W. 1998.** Soil microbial diversity and community structure under wheat as influenced by tillage and crop rotation. *Soil Biol. Biochem.* **30**: 1733–1741.

**Martens-Habbena, W., Berube, P. M., Urakawa, H., de la Torre, J. R. and Stahl, D. A. 2009.** Ammonia oxidation kinetics determine niche separation of nitrifying archaea and bacteria. *Nature* **461**: 976–979.

**Metting, F. B. 1993.** Structure and physiological ecology of soil microbial communities. Pages 3–24 in F. B. Metting, ed. *Soil microbial ecology – Application in agricultural and environmental management*. Marcel Dekker, New York, NY.

**Muhling, M., Woolven-Allen, J., Murrell, J. C. and Joint, I. 2008.** Improved group-specific PCR primers for denaturing gradient gel electrophoresis analysis of the genetic diversity of complex microbial communities. *ISME J.* **2**: 379–392.

**Niederberger, T. D., Perreault, N., Tille, S., Sherwood, L. B., Lacrampe-Couloume, G., Andersen, D., Greer, C. W., Pollard, W. and Whyte, L. G. 2010.** Microbial characterization of a subzero, hypersaline methane seep in the Canadian high arctic. *ISME J.* **4**: 784–798.

**Patricia, B. B. and Gordon, V. W. 2006.** Protist genetic diversity in the acidic hydrothermal environments of Lassen Volcanic National Park, USA. *J. Eukaryotic Microbiol.* **53**: 420–431.

**Paustian, K., Collins, H. P. and Paul, E. A. 1997.** Management controls on soil organic carbon. Pages 15–51 in E. A. Paul, K. Paustian, E. T. Elliott, and C. V. Cole, eds. *Soil organic matter in temperate agroecosystems: Long-term experiments in North America*. CRC Press, Boca Raton, FL.

**Peixoto, R., Coutinho, H., Madari, B., Machado, P., Rumjanek, N., Van Elsas, J., Seldin, L. and Rosado, A. 2006.** Soil aggregation and bacterial community structure as affected by tillage and cover cropping in the Brazilian Cerrados. *Soil Tillage Res.* **90**: 16–28.

- Perreault, N. N., Andersen, D. T., Pollard, W. H., Greer, C. W. and Whyte, L. G. 2007. Characterization of the prokaryotic diversity in cold saline perennial springs of the Canadian high arctic. *Appl. Environ. Microbiol.* **73**: 1532–1543.
- Quaiser, A., Ochsenreiter, T., Lanz, C., Schuster, S. C., Treusch, A. H. and Schleper, C. 2003. Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. *Mol. Microbiol.* **50**: 563–575.
- Roesch, L. F. W., Fulthorpe, R. R., Rive, A., Casella, G., Hadwin, A. K. M., Kent, A. D., Daroub, S. H., Camargo, F. A. O., Farmerie, W. G. and Triplett, E. W. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J.* **1**: 283–290.
- SAS Institute, Inc. 2009. The SAS system for windows, release 9.2. SAS Institute, Inc., Campus Drive, Cary, NC.
- Schloss, P. D. and Handelsman, J. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl. Environ. Microbiol.* **71**: 1501–1506.
- Spang, A., Hatzepichler, R., Brochier-Armanet, C., Rattei, T., Tischler, P., Spieck, E., Streit, W., Stahl, D. A., Wagner, M. and Schleper, C. 2010. Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. *Trends Microbiol.* **18**: 331–340.
- Spedding, T. A., Hamel, C., Mehuys, G. R. and Madramootoo, C. A. 2004. Soil microbial dynamics in maize-growing soil under different tillage and residue management systems. *Soil Biol. Biochem.* **36**: 499–512.
- Statistics Canada. 2011a. CANSIM Table 001-0010 and 001-0018: Estimated areas, yield, production, average farm price of principal field crops. Statistics Canada, Ottawa, ON.
- Statistics Canada. 2011b. CANSIM Table 001-0018: Estimated areas, yield, production, average farm price and total farm value of selected principal field crops. Statistics Canada, Ottawa, ON.
- Steven, B., Briggs, G., McKay, C. P., Pollard, W. H., Greer, C. W. and Whyte, L. G. 2007. Characterization of the microbial diversity in a permafrost sample from the Canadian high arctic using culture-dependent and culture-independent methods. *FEMS Microbiol. Ecol.* **59**: 513–523.
- Steven, B., Pollard, W. H., Greer, C. W. and Whyte, L. G. 2008. Microbial diversity and activity through a permafrost/ground ice core profile from the Canadian high arctic. *Environ. Microbiol.* **10**: 3388–3403.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**: 2731–2739.
- Tourna, M., Stieglmeier, M., Spang, A., Könneke, M., Schintlmeister, A., Ulrich, T., Engel, M., Schloter, M., Wagner, A., Richter, A. and Schleper, C. 2011. *Nitrososphaera viennensis*, an ammonia oxidizing archaeon from soil. *PNAS* **108**: 8420–8425.
- Upchurch, R., Chiu, C., Everett, K., Dyszynski, G., Coleman, D. and Whitman, W. 2008. Differences in the composition and diversity of bacterial communities from agricultural and forest soils. *Soil Biol. Biochem.* **40**: 1294–1305.
- Valinsky, L., Della Vedova, G., Scupham, A. J., Alvey, S., Figueroa, A., Yin, B., Hartin, R. J., Chrobak, M., Crowley, D. E., Jiang, T. and Borneman, J. 2002. Analysis of bacterial community composition by oligonucleotide fingerprinting of rRNA genes. *Appl. Environ. Microbiol.* **68**: 3243–3250.
- Van Groenigen, K. J., Bloem, J., Bââth, E., Boeckx, P., Rousk, J., Bode, S., Forristal, D. and Jones, M. B. 2010. Abundance, production and stabilization of microbial biomass under conventional and reduced tillage. *Soil Biol. Biochem.* **42**: 48–55.
- Voroney, R. P., Brookes, P. C. and Beyaert, R. P. 2008. Soil microbial biomass C, N, P and S. Pages 637–651 in M. R. Carter and E. G. Gregorich, eds. *Soil sampling and methods of analysis*. 2nd ed. CRC Press, Boca Raton, FL.
- Vyn, T. J. and Raimbault, B. A. 1993. Long-term effect of five tillage systems on corn response and soil structure. *Agron. J.* **85**: 1074–1079.
- Whalen, J. K. and Sampedro, L. 2010. *Soil ecology and management*. CABI International, Oxfordshire, UK. 320 pp.
- Whyte, L. G. and Greer, C. W. 2005. Molecular techniques for monitoring and assessing soil bioremediation. R. Margesin and F. Schinner, eds. Pages 201–231 in *Manual for soil analysis – Monitoring and assessing soil bioremediation*. Springer Verlag, Berlin, Germany.
- Yin, C., Jones, K. L., Peterson, D. E., Garrett, K. A. and Hulbert, S. H. 2010. Members of soil bacterial communities sensitive to tillage and crop rotation. *Soil Biol. Biochem.* **42**: 2111–2118.
- Zhang, H., Parameswaran, P., Badalamenti, J., Rittmann, B. E. and Krajmalnik-Brown, R. 2011. Integrating high-throughput pyrosequencing and quantitative real-time PCR to analyze complex microbial communities. *Methods Mol. Biol.* **733**: 107–128.



# THRESHOLDS FOR IRRIGATION MANAGEMENT OF PROCESSING TOMATOES USING SOIL MOISTURE SENSORS IN SOUTHWESTERN ONTARIO

F. Jaria, C. A. Madramootoo

**ABSTRACT.** *Processing tomato (Lycopersicon esculentum Mill.) is an economically important vegetable crop in southwestern Ontario. Processing tomato (cultivar H9553) fruit yield and quality were evaluated in field experiments in southwestern Ontario over a three-year period (2008-2010). A split-plot randomized complete block design with four blocks was used in 2008 and 2010. Irrigation types (buried and surface drip) served as the main plots, while four moisture depletion levels constituted the split plots. In 2009, a 2x4 factorial complete randomized block design with four blocks was used, with the same two factors. The moisture treatments represented the lower soil moisture triggers, which initiated irrigation scheduling. Irrigation was terminated for each treatment when field capacity was reached. Continuous soil moisture status over the growing season was monitored with a combination of volumetric and tensiometric sensors. Seven fruit quality parameters were monitored: fruit weight, color, pH, size, firmness, Brix yield, and soluble solids. In each year, the most stressed treatment produced the highest soluble solids (6.0, 4.8, and 5.2 °Brix for 2008, 2009, and 2010, respectively). Total and marketable fruit yields ranged from 91.9 to 121.1 Mg ha<sup>-1</sup> and from 91.4 and 119.7 kg ha<sup>-1</sup>, respectively. Statistical significance was obtained among treatments and irrigation types in 2008 only. Irrigation water use efficiency was also not statistically significant over the three years. Seasonal irrigation depth ranged from 58 to 196 mm, and statistical significance among the moisture treatments was obtained in 2008 and 2010.*

**Keywords.** *FDR, Irrigation scheduling, Irrigation thresholds, TDR.*

Agriculture is a key driver for the Canadian economy, providing one in seven jobs within the country. The agri-food sector accounts for 8.3% of Canada's gross domestic product, USD \$26.5 billion of which comes from exports and employing nearly 2.1 million persons (AAFC, 2006). Canadian vegetable growers reported sales of USD \$659 million in 2010, with two provinces (Ontario and Quebec) accounting for more than 80% of the vegetable sales (Statistics Canada, 2011). Vegetable and dry bean production is a critical part of the food and agriculture industry in Ontario. Virtually all of the tomatoes grown in Canada for processing are produced in Ontario, with the counties of Essex and Kent being the main producing areas. In 2008, 0.62 million tons of processing tomatoes were produced by 150 growers, generating over USD \$60.5 million (OHCRSC, 2006). In most years, rainfall during the growing season is insufficient to attain optimum production (Warner et al., 2007). Tan et al. (2003) noted that, through the 1990s,

rainfall during the growing season decreased by about 25 mm year<sup>-1</sup>. The 30-year (1971 to 2000) climate normal rainfall for Windsor and London averaged 254.3 and 251.3 mm, respectively. Over the growing season, an average cultivar requires 400 mm of water (LeBoeuf et al., 2007). Thus, intensive tomato production in these two counties necessitates the use of supplemental irrigation to offset the deficiencies in rainfall to maintain high levels of production (Warner et al., 2007).

There is increasing pressure for more judicious utilization of limited water resources to reduce negative environmental impacts. Shock et al. (2001) identified economic competition in marketing produce, competition for water, and political pressure as the three forces to minimize off-site impacts of irrigation-induced runoff and leaching. It is desirable to optimize crop yield and quality under the constraints of reducing water use and increasing the efficiency of the use of agricultural chemicals.

Irrigation scheduling is a technique for timely and accurately application of water to a crop and is key to conserving water, improving irrigation performance, and ensuring the sustainability of irrigated agriculture (Thompson et al., 2007b). Several irrigation scheduling methods based on water budget, soil, and plant indicators have been used for different crops, with the water budget method probably the most widely used technique (Fareres et al., 2003). However, over the past decade, a new generation of soil moisture sensors based on electrical properties, such as resistance, capacitance, and time

---

Submitted for review in June 2012 as manuscript number SW 9792; approved for publication by the Soil & Water Division of ASABE in January 2013.

The authors are **Felix Jaria**, ASABE Member, Graduate Student, and **Chandra A. Madramootoo**, Professor and Dean of Agricultural and Environmental Sciences, Department of Bioresource Engineering, McDonald Campus of McGill University, Ste. Anne de Bellevue, Quebec, Canada. **Corresponding author:** Felix Jaria, Department of Bioresource Engineering, McDonald Campus of McGill University, 21111 Lakeshore Rd, Ste. Anne de Bellevue, Quebec, Canada H9X 3V9; e-mail: felix.jaria@mail.mcgill.ca; phone: 514-398-8785.

domain reflectometry, has been developed (Fares et al., 2006; Fereres et al., 2003). These sensors measure either soil matric potential or volumetric soil water content (Thompson et al., 2007b). Soil moisture sensors facilitate frequent but low-volume irrigation applications, which have been found to be superior to the traditional scheduling of fewer, large-volume applications. These devices have been used extensively for efficient irrigation and nutrient management in different crops (Fares and Alva, 2000; Lukangu et al., 1999; Thompson et al., 2007b).

Soil moisture sensors can be used as a standalone method to effect irrigation scheduling or in conjunction with other methods, such as the FAO and water budget methods (Thompson et al., 2007a, 2007b). However, optimal irrigation scheduling using soil moisture, whether soil matric potential or volumetric soil water content, necessitates accurate threshold values or indices for individual crops in a given agricultural system (Lukangu et al., 1999). The upper and lower thresholds are described as the fill and refill points, respectively, with the fill point corresponding to field capacity. The refill point is the soil moisture content below which crop growth is measurably decreased or where a crop begins to experience water stress (Campbell and Campbell, 1982).

For volumetric soil moisture content sensors, the available water content (AWC) is often used to determine the trigger or threshold for irrigation management (Thompson et al., 2007b). Since the AWC is the moisture available to the plant between field capacity and permanent wilting point, a management allowable depletion (MAD) ranging between 20% and 40% of AWC is often used as the refill threshold for different crops. FAO 56 provides guidelines on these levels and recommends a depletion of 40% for tomato (Allen et al., 1998). Hartz et al. (2005) added that tomatoes can tolerate a depletion of 20% to 30% of available soil moisture in the active root zone without having yield losses. It is also possible to establish threshold values for soil moisture sensors as a percentage of field capacity instead of the MAD, since the two are related. Shock et al. (2007) added that irrigation thresholds should be determined for site-specific conditions to account for variability in climate, soils, crop cultivars, and irrigation systems.

A range of threshold/refill values has been used for tension-based soil moisture sensors (Shock and Wang, 2011). Haise and Hagan (1967) used refill points of -60 to -70 kPa for high and low evaporative demand conditions for cabbages. Stanley and Maynard (1990) recommended soil water potential levels in the -10 and -30 kPa range for vegetables grown under irrigation with high and low evaporative demand, respectively. Thompson et al. (2007b) reported soil matric potential threshold values of -35 kPa for melon and -38 to -58 kPa for tomatoes. Marouelli and Silva (2007) tested tension threshold values between -5 and -120 kPa for processing tomatoes and found that soil moisture tension thresholds of -35, -12, and -15 kPa produced the highest yields for the vegetative, fruit development, and maturation growth stages, respectively.

Numerous studies have found that irrigation substantially increased fruit yield of processing tomato (Hanson et al., 2006; Patanè and Cosentino, 2010). Warner et al. (2007) obtained processing tomato yields ranging from 126.1 to 181.5 Mg ha<sup>-1</sup> under different irrigation application rates as a function of ET<sub>c</sub> (0.5 to 1.2 ET<sub>c</sub>) for Harrow, Ontario. At the University of California, Johnstone et al. (2005) carried out drip irrigation experiments between 2000 and 2002 on processing tomatoes in loam soils, as a function of reference evapotranspiration, and found yields ranging from 78 to 125 Mg ha<sup>-1</sup>. Drip irrigation experiments using varying amounts of potassium (from 0 to 600 kg ha<sup>-1</sup>) produced total yields for processing tomatoes (cultivar H9478) ranging from 86.6 to 92.5 Mg ha<sup>-1</sup> at Harrow, Ontario (Liu et al., 2011).

In the humid climate of southwestern Ontario, determining the optimum amount of water to apply to accurately meet crop water requirements, especially after a rainfall event, can be a challenge. To this end, the objective of this work was to use a combination of tension and volumetric based continuous soil moisture monitoring sensors as standalone devices to determine the optimum trigger point (refill) and to effect irrigation scheduling of field processing tomatoes grown in loamy sand soils.

## MATERIALS AND METHODS

### EXPERIMENTAL LOCATION

A three-year (2008 to 2010) field experiment was conducted during the summer months of May through September on a commercial farm in Leamington, in southwestern Ontario. The climate is classified as humid, with hot summers complemented by dry and cold winters. Average annual temperature and precipitation are approximately 9.5°C and 750 mm, respectively (Tan and Reynolds, 2003). The growing season for field processing tomatoes extends from mid-May to September, with average growing season maximum and minimum daily temperatures 24.9°C and 15.0°C, respectively, and average seasonal rainfall of approximately 261.1 mm. The rainfall is typically spread throughout the year, with no predominant rainy months. The dominant soil within the production zone (0 to 30 cm) is loamy sand (86% sand, 8% silt, and 6% clay) with an average bulk density of 1450 kg m<sup>-3</sup>. Field soil water capacity ranged between 20% and 25% by volume. Chemical properties of the soil (0 to 30 cm) are provided in table 1. Agronomic soil test P and NO<sub>3</sub>-N were determined using the Olsen P procedure (0.5 M NaHCO<sub>3</sub>, pH 8.5; Olsen et al., 1954) and the 2.0 M KCl procedure (Keeney and Nelson, 1982), respectively.

**Table 1. 2008-2010 pre-planting soil properties at the 0 to 30 cm depth at the experimental site.**

Soil Parameter	2008	2009	2010
NO <sub>3</sub> -N (ppm)	52	101	37
Available P (ppm)	144	121	154
Potassium (ppm)	243	219	191
pH	7.3	7.0	7.0
Organic matter (%)	2.1	2.9	2.9

**Table 2. Experimental design over the three years (RCBD = randomized complete block design; FC = field capacity).**

Year	Experimental Design	Factor 1: Irrigation Type		Factor 2: Moisture Level			Tension base	Blocks
2008	Split-plot	Surface drip	Buried drip	60%	70%	80%	(-30 kPa)	4
	RCBD	irrigation	irrigation	FC	FC	FC		
2009	Factorial	Surface drip	Buried drip	74%	82%	91%	(-30 kPa)	4
	RCBD	irrigation	irrigation	FC	FC	FC		
2010	Split-plot	Surface drip	Buried drip	55%	70%	85%	(-30 kPa)	4
	RCBD	irrigation	irrigation	FC	FC	FC		

## EXPERIMENTAL DESIGN

Table 2 provides a summary of the experimental design over the three years. A split-plot randomized complete block design (RCBD) was used in 2008 and 2010, and a 2×4 factorial RCBD was used in 2009. The split-plot design involved two experimental factors. The irrigation types (buried and surface drip irrigation) and moisture levels (three moisture levels and a tension treatment) were assigned to the whole plot (main plot) and split plots, respectively. The factorial experimental design in 2009 had the same two factors. The volumetric moisture treatments were expressed as a fraction of field capacity (FC) and represented the depletion that the soil moisture reached to initiate irrigation scheduling. In 2009, the volumetric moisture treatments were changed from 60%, 70%, and 80% of FC to 74%, 82%, and 91% of FC. This was done to examine the effects of a less stressed irrigation scheduling program. In 2010, the moisture treatments were again changed to 55%, 70%, and 80% of FC. This was done to increase the range between moisture treatments, making it more practical for monitoring. The change in the experimental design from split-plot RCBD to a factorial design was due to the fact that the parameters involved in the experiment fitted both models, and it was felt opportune to change the model in one of the three years.

The experiment for each year had four blocks, with each block having eight plots, for a total of 32 plots (16 with buried drip irrigation and 16 with surface drip irrigation). Due to the annual crop rotation, the experimental site on the farm was changed each year. In 2008, the blocks were oriented across the beds. Each plot comprised adjacent twin beds (2 beds × 1.5 m × 7.5 m) with an area of 22.5 m<sup>2</sup> per plot and located between guard beds (1.5 m × 7.5 m) on either side. There was a 1.5 m buffer between blocks. In 2009 and 2010, the blocks were oriented along the beds. Each of the plots (treatments) per block comprised a single bed (1.5 m × 8.0 m) with an area of 12 m<sup>2</sup> per plot and located between guard beds (1.5 m × 8 m) on either side. The guard beds formed the separation between blocks, and there was a 1.5 m buffer between plots in the blocks.

One drip line (irrigation tape, Streamline 636 006 F, Netafim Irrigation, Inc., Fresno, Cal.) was aligned along the surface of each twin-row bed for the surface irrigated plots. For the buried irrigated plots, the drip lines were installed to a depth of 20 cm (in 2008) and 15 cm (in 2009 and 2010). The inline emitters were spaced 30 cm apart, with a flow rate of at 0.46 L h<sup>-1</sup> at 55 kPa in 2008 and 0.68 L h<sup>-1</sup> at 62 kPa in 2009 and 2010, providing a uniform soil wetting pattern. Each plot had the same number of emitters. The volume of water applied during each irrigation event to each plot was determined as the product of the irrigation duration and the flow rate per plot at the requisite water

pressure. The equivalent irrigation depth was determined as the quotient of the irrigation volume and effective wetted area.

## CROPPING DETAILS

Processing tomato (*Lycopersicon esculentum* Mill., cultivar Heinz H9553) was grown in the study area during the three years. Seedlings (42 days old) were transplanted in soil with water content near field capacity in the top 30 cm. Transplant dates were 29 May 2008, 25 May 2009, and 15 May 2010. The crop was harvested after 105 days in 2008 (on 10 Sept.), after 112 days in 2009 (on 14 Sept.), and after 101 days in 2010 (on 24 Aug.). The seedlings were spaced 42 cm apart within the two rows, and the rows were 50 cm apart. Each set of twin rows was centered on a 1.5 m wide raised bed. The plant density was 31,746 plants ha<sup>-1</sup>.

## SOIL MOISTURE SENSORS, DATA COLLECTION, AND SENSOR CALIBRATION

Three types of soil moisture sensors were installed in the field for continuous data collection: a time domain reflectometer (TDR) (CS625 water content reflectometer, Campbell Scientific Inc., Logan, Utah), a frequency domain reflectometer (FDR) (EnviroSMART, Sentek Sensor Technologies, Stepney, Australia), and an electronic tensiometer (Irrolis Sense Tx, Hortau, Inc., Saint-Romuald, Quebec, Canada).

The TDRs were installed with the aid of an insertion guide. The procedure for installing the FDR access tubes and the FDR sensors onto the probe guide is explained in the Sentek manual (Sentek, 2003). The tensiometers were installed in the conventional manner. For irrigation scheduling purposes, the critical depths at which the three soil moisture sensors were monitored were 0 to 30 cm for the TDRs, 20 cm for the FDRs (which effectively measured the soil moisture content over the 5 to 25 cm depth), and 15 cm for the tensiometers. All sensors (or access tubes in the case of the FDRs) were installed 10 cm away from the centrally aligned drip line and 10 cm away from the nearest emitter to ensure consistency in data collection.

Some changes occurred during the 2009 and 2010 experimental seasons. In 2009, the critical monitoring depth for the TDRs was 5 to 25 cm. The top 5 cm of soil was removed (at the site of installation), and the sensor was installed at a 33.6° angle to the vertical plane. In 2010, only TDRs and tensiometers were installed in the experimental plots. The monitoring depth for the TDRs was reverted back to 0 to 30 cm. Also in 2010, an upgraded version of the Irrolis Sense Tx electronic tensiometer, called the Irrolis MultiSense Tx3 probe, was used.

All devices were equipped with wireless communication



to transmit data from the field to an onsite computer. All volumetric sensors were connected to 12 solar-powered data loggers (model CR205/6, Campbell Scientific, Inc.). The data were scanned every 5 min and recorded every 15 min, hourly and daily. The data were retrieved from the CR205/6 using a computer and LoggerNet software (Campbell Scientific, Inc.). The electronic tensiometer data were transmitted in real-time by wireless signals to the onsite desktop computer, which was equipped with the requisite proprietary hardware and software. Meteorological data were collected from an on-site weather station from 1 May to 31 August in 2008, 2009, and 2010. The weather parameters measured included temperature, rainfall, relative humidity, solar radiation, and wind speed.

Generic calibration curves for the TDR and FDR sensors were developed for site-specific conditions over the rooting depth of the crop (0 to 30 cm) against measured volumetric data. The tensiometers were not calibrated. *In situ* field capacity measurements were determined using the combined procedures outlined by Peters (1965) and Ratliff et al. (1983).

### IRRIGATION SCHEDULING

From the calibration curves for the TDR and FDR sensors, the upper and lower volumetric water content and sensor threshold values (in  $\mu\text{s}$  and SFU for the TDR and FDR, respectively) were determined. The predetermined upper and lower threshold values for the tension-based treatment were -10 and -30 kPa, respectively. For all the treatments, the upper threshold value was FC. The soil moisture sensors for all 32 plots were continuously monitored at a central location. When the soil moisture content for each plot (buried and surface drip irrigated) was depleted to its requisite moisture treatment threshold value, irrigation was initiated. Irrigation was terminated when the upper trigger (FC) moisture content was reached. The irrigation scheduling process for each plot was done throughout the growing season. The volume of irrigation and the equivalent irrigation depth applied during each irrigation event were determined for each plot throughout the irrigation scheduling program. The irrigation scheduling program was implemented throughout the irrigation season, which began on 7 July, 29 June, and 21 June and was terminated on 1 September, 31 August, and 15 August in 2008, 2009, and 2010, respectively. In addition to the preplanting and side-dressing nutrient applications, all plots were fertigated simultaneously during each of the three years and for the same duration.

### CROP YIELD, CROP QUALITY, AND IWUE

The fruits were harvested approximately ten days after spraying ethrel. To evaluate fruit yield and quality, all the fruits were harvested from six plants (2008 and 2010) or from four plants (2009) in each sub-plot. Fruits were categorized into red, green, or cull and weighed to determine total and marketable yields. Marketable yield was obtained by subtracting the weight of the culled fruits from the total yield. A random sample from each plot was tested for soluble sugar content, pH, firmness, fruit size, and color. Total solids

were determined as the product of soluble solids and harvestable yields. From each plot, a random sample of approximately 50 marketable fruits was used to determine average fruit weight (Warner et al., 2007). A random sample of approximately 20 marketable fruits from each plot was tested for firmness with the use of a penetrometer (model FT 0110, Facchini, Alfonsine, Italy). The penetrometer was equipped with a cylindrical pin, 5 cm long with a 2 mm diameter flat end. The average of two firmness measures was taken on each fruit at opposite sides of the equatorial zone. A random sample of 20 red ripe fruits from each plot was washed, skinned, deseeded, and made into a pulp. The soluble solids ( $^{\circ}\text{Brix}$ ) and pH were measured in the homogenized juice using a digital refractometer and pH meter, respectively. Irrigation water use efficiency (IWUE) was determined by the ratio of the marketable yield ( $\text{kg ha}^{-1}$ ) and the total seasonal irrigation volume ( $\text{m}^3 \text{ha}^{-1}$ ), including rainfall, and expressed as  $\text{kg m}^{-3}$  (Howell, 2001).

### STATISTICAL ANALYSIS

Statistical analysis was performed on individual years of data. In both experimental design models, the blocks were considered random effects, while the irrigation types and the moisture levels (treatments) were fixed-effects parameters. Statistical analyses were performed using the PROC MIXED procedure of SAS (SAS, 2007), designed to fit mixed-effect models. Analysis of fruit yield, fruit quality, and irrigation parameters was done. Differences at  $p < 0.05$  were considered statistically significant. Least squares means of fixed-effects parameters were pairwise compared at  $p < 0.05$ . Model assumptions (normal distribution, and consistent variance of error terms) were verified prior to carrying out the analysis.

## RESULTS AND DISCUSSION

### WEATHER PATTERNS AND CLIMATIC CONDITIONS

The weather data are summarized in table 3. There was no significant difference in the rainfall totals (3.5 mm difference) over the monitoring period in 2008 and 2009; however, the distribution was very different. In 2008, May and June had higher rainfall than July and August, while the opposite was true in 2009. A comparison with the 30-year climate normal (1971 to 2000) indicated that the summer months of 2008 and 2009 received 15% and 16% lower than normal precipitation, respectively. The 2010 monthly rainfall exceeded the normal rainfall for the growing season, with the exception of August. The 2010 total rainfall also exceeded the 30-year average by 42.7%. The average over the three years (May to August) was 261.2 mm. Zhang et al. (2010) also reported average rainfall of 291.8 mm over the growing period (May to Sept) in Harrow, Ontario, Canada.

Air temperature gradually increased from about  $15^{\circ}\text{C}$  at the beginning of May to about  $30^{\circ}\text{C}$  between June and mid-August, after which there was a gradual decrease toward the end of August for all three years. Just prior to planting in May 2010, there was a four-day period in which the

**Table 3. Rainfall comparison to 30-year normal (1971-2000).**

Month	Climate Normal (1971 to 2000) Rainfall (mm)		Actual Rainfall (mm)						Average Temperature (°C)		
	London	Windsor	Leamington			Difference from Normal (%, relative to Windsor)			Leamington		
			2008	2009	2010	2008	2009	2010	2008	2009	2010
May	82.6	80.7	66.4	12.5	114.7	-18.7	-84.7	72.7	12.5	14.7	15.4
June	86.8	89.8	108.1	65.0	91.4	22.4	-26.4	1.8	20.6	18.8	21.1
July	82.2	81.8	30.6	34.4	135.5	-62.7	-58	57.8	22.7	20.1	24.4
August	85.3	79.7	9.5	98.6	16.9	-88.5	19.5	78.8	20.7	21.0	23.4
Total	254.3	251.3	214.6	210.5	358.5	-15.1	-16.7	42.7			

temperatures dipped below 15°C. Although the trends were similar in the three years, the average monthly, maximum, and minimum air temperatures showed some variation. Jones (2007) highlighted the fact that tomato is a day length neutral plant under conditions of short or long days, requiring optimum mean daily temperature of 18.5°C to 25°C for growth, with night temperatures between 18°C and 21°C. At the time of planting, these conditions were met.

#### FIELD CAPACITY AND AVAILABLE WATER CONTENT

The FC and permanent wilting point (PWP) for 2008 to 2010 were 20% and 9%, 22% and 9%, and 25% and 10% volumetric water content (VWC), respectively. The varying FC values obtained may be attributed to the changing organic matter content over the years due to crop rotation and land preparation. Hudson (1994) noted that within each textural group, as organic matter (OM) content increased, the volume of water held at field capacity increased at a much greater rate than that held at PWP. He further added that 1% to 6% OM by weight was equivalent to approximately 5% to 25% by volume and hence can have a significant effect on available water content (AWC). The

**Table 4. Treatments expressed as depletion of AWC.**

Year	Treatment	AWC Depletion (% VWC)
2008	80% FC	36
	70% FC	54
	60% FC	73
	-30 kPa	24
2009	91% FC	15
	82% FC	30
	74% FC	45
	-30 kPa	22
2010	85% FC	25
	70% FC	58
	55% FC	75
	-30 kPa	20

AWC ranged between 10% and 15%  $\theta_v$  and was within the range (9% to 15%) of values for the soil type, as provided by Schwab et al. (1993). Table 4 summarizes the FC and tension treatments over the three years as fractions of AWC. Doorenbos and Pruitt (1977) recommended a management allowable deficit (MAD) for tomatoes of 30% to 40% of AWC to facilitate maximum yields. In 2008 and 2010, two of the four treatments were within that range; in 2009, three of the four treatments were within that range.

#### APPLIED WATER IN 2008

The irrigation scheduling summary and statistical results are shown in tables 5 and 6. In 2008, irrigation duration was not statistically significant among moisture treatments, irrigation types, or the interaction between moisture treatment and irrigation type. However, statistical significance for irrigation events, equivalent depths, and irrigation volumes was obtained among the moisture treatments, but not among irrigation types nor for the interaction between irrigation type and moisture treatment.

The -30 kPa treatment represented the least stress treatment, corresponding to a soil moisture content of approximately 88% FC or 24% depletion in AWC. It invariably reached the trigger point more frequently than the other moisture treatments and therefore received the most irrigation. Further, the tension treatment represented a point measurement of moisture content at the 15 cm depth, as compared to the TDRs and FDRs, which reflected the integrated moisture content over the 0 to 30 cm and 15 to 25 cm depths, respectively. This shallower depth also meant that the tension treatment reached the trigger point more frequently.

The 70% FC treatment was irrigated for a longer duration, received more water, and was irrigated more frequently than the 80% FC treatment. Two factors contributed to this discrepancy. First, one of the eight 80%

**Table 5. Irrigation scheduling summary per treatment.**

	Treatment	Duration (h)	Volume Applied (L)	Equivalent Depth (mm)	Effective Rainfall (mm)	Total Irrigation Events
2008 Season Averages	60% FC	59.8	1464	114	167.2	12
	70% FC	81.9	2006	156	167.2	20
	80% FC	53.7	1317	103	167.2	17
	-30 kPa	88.1	2160	168	167.2	30
2009 Season Averages	74% FC	32.9	582	85	186	11
	82% FC	66.8	1182	173	186	16
	91% FC	73.6	1302	190	186	23
	-30 kPa	58.6	1036	151	186	27
2010 Season Averages	55% FC	22.5	399	58	294.2	4
	70% FC	39.9	705	103	294.2	7
	85% FC	70.0	1238	181	294.2	15
	-30 kPa	75.9	1342	196	294.2	18



**Table 6. Statistical results of fixed effect parameters: Irrigation parameters (2008-2010).<sup>[a]</sup>**

Irrigation Parameter	2008			2009			2010		
	Moisture Treatment	Irrigation Type	MT × IT	Moisture Treatment	Irrigation Type	MT × IT	Moisture Treatment	Irrigation Type	MT × IT
	(MT)	(IT)		(MT)	(IT)		(MT)	(IT)	
Irrigation duration	NS	NS	NS	NS	NS	NS	*	NS	NS
Total irrigation events	*	NS	NS	*	NS	NS	*	NS	NS
Irrigation equivalent depth	*	NS	NS	NS	NS	NS	*	NS	NS
Irrigation volume	*	NS	NS	NS	NS	NS	*	NS	NS

<sup>[a]</sup> NS = not significant; asterisk (\*) indicates significance at  $p \leq 0.05$ .

FC plots recorded unusually high moisture content throughout the growing season and was only irrigated four times, as compared to an average of 17 times for the 80% FC treatment. The consistently high moisture content in that plot was due to a depression in the field, which allowed lateral movement of soil moisture to accumulate in the vicinity of the plot. Secondly, one of the 70% FC plots was irrigated 14 times more than the average 70% FC plot, while another was irrigated for 50 h more than the average 70% FC treatment.

Although the 80% FC treatment was irrigated more regularly than the 60% FC treatment, the average seasonal irrigation depth was higher for 60% FC than for 80% FC. This was due to a combination of factors. One of the factors was highlighted above, with reference to the high moisture content of one of the 80% FC plots. Secondly, one of the 60% FC plots was irrigated more frequently than the average 80% FC treatment. Thirdly, the fact that the 60% FC treatment represented a wider threshold range meant that the average duration of application per irrigation event for the 60% FC treatment was longer than for the 80% FC treatment.

#### APPLIED WATER IN 2009 AND 2010

In 2009, the treatments reflected a less stressed irrigation scheduling program than in the 2008 experiment, resulting in statistical significance only for the irrigation events among the moisture treatments. There was no statistical significance among the irrigation types nor for the interaction between irrigation type and moisture treatment. The average seasonal irrigation volume per plot was less than in 2008 because the plot area was reduced by approximately 47%. The equivalent depth of irrigation applied increased with the increase of the FC treatments (from 74% FC to 91% FC) and decreased for the -30 kPa treatment. Despite higher effective rainfall in 2009 and less stressed irrigation treatments, the equivalent irrigation depths for two of the FC treatments were higher in 2009 than in 2008. In 2008, the rainfall was concentrated during the first two months of the growing season; as a result, more effective use was made of the rainfall. Therefore, the need for irrigation during the early part of the season was minimal. In 2009, the rainfall was concentrated toward the end of the growing season, and there was therefore more need to irrigate during the early season. Most of the rainfall was in August; by then, the crop was close to harvesting, and the need for irrigation was minimal. The August rainfall therefore was not used very effectively.

In 2010, the FC treatments were changed to reflect a larger range between treatments, which contributed to

statistical significance between irrigation duration, irrigation events, equivalent depth, and irrigation volume among the moisture treatments. There was no statistical significance among the irrigation types and the interaction between irrigation type and moisture treatment. There was a consistent trend among the treatments, with the most stressed moisture treatment reaching the trigger point the fewest number of times and therefore receiving the least seasonal irrigation and the shortest irrigation duration. The reciprocal was obtained for the least stressed treatment. The FC treatments were monitored at 0 to 30 cm depth, and the tension treatment was monitored at 15 cm depth. The shallower soil depth at which the tension treatment was measured would account for it reaching the trigger point more frequently. The distribution and total depth of the effective rainfall over the growing season would have contributed to the fewer overall irrigation events for all the treatments.

Warner et al. (2007) reported seasonal irrigation depths ranging from 58 to 267.6 mm during a three-year experiment in Harrow, Ontario (average rainfall of 247.1 for June to August) for different surface drip irrigation treatments for processing tomatoes in Granby sandy loam. Machado and Oliveira (2005) also reported values of 243.1 to 560.9 mm for subsurface drip irrigation in Coruche, Portugal, in sandy soils, with rainfall over the growing season amounting to 76.1 mm. In both cases, irrigation was applied as a function of crop evapotranspiration.

#### FRUIT YIELDS IN 2008

Tables 7 and 8 summarize yield results and yield statistics. In 2008, the average fruit yield for the four treatments ranged from 91.9 to 121 Mg ha<sup>-1</sup>. The highest and lowest average yields were from the 70% FC and 60% FC treatments, respectively. There was a direct relationship between seasonal irrigation depth and crop yield (fig. 1). As a result, statistical significance was obtained for total and marketable yields among moisture treatments as well as irrigation types. Pairwise comparisons between moisture treatments revealed statistical significance only between 60% FC and 70% FC, between 60% FC and -30 kPa, and between 70% FC and 80% FC, but not between 60% FC and 80% FC, nor between -30 kPa and 80% FC. No significant difference was found in the interaction between moisture treatment and irrigation type.

The 70% FC treatment represented a depletion of the AWC of approximately 54%, which was substantially lower than the 30% to 40% AWC recommended by Doorenbos and Pruitt (1977). It was therefore surprising that the 70% FC treatment produced the highest yield. A

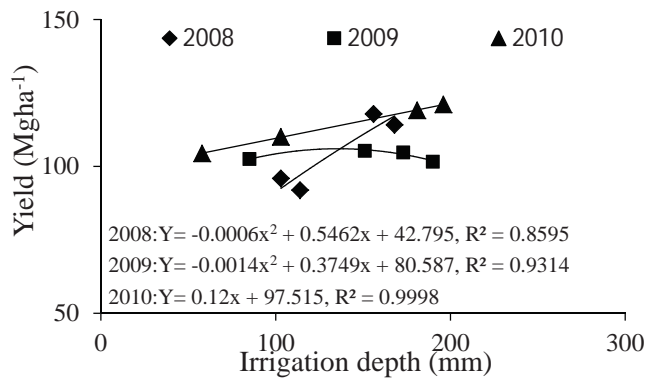
**Table 7. Fruit yields per treatment for 2008-2010 (SE = standard error).**

Year	Treatment	Total Yield (Mg ha <sup>-1</sup> )		Marketable Yield (Mg ha <sup>-1</sup> )		Red Fruit Yield (Mg ha <sup>-1</sup> )	
		Mean	SE	Mean	SE	Mean	SE
2008	60% FC	91.9	8.6	91.4	8.6	80.3	8.6
	70% FC	117.8	9.5	117.0	9.3	107.2	9.0
	80% FC	95.9	8.7	95.3	8.6	88.6	8.5
	-30 kPa	114.1	5.2	114.0	5.2	107.0	4.1
2009	74% FC	102.5	3.2	99.8	2.7	98.5	2.8
	82% FC	104.7	4.2	102.1	4.4	100.9	4.4
	91% FC	101.6	3.3	98.0	3.3	97.8	3.2
	-30 kPa	105.3	3.3	102.9	3.2	102.3	3.3
2010	55% FC	104.4	5.4	102.0	5.6	90.6	8.6
	70% FC	110.0	6.9	107.1	6.9	99.7	7.1
	85% FC	119.1	5.0	115.9	5.4	110.8	5.8
	-30 kPa	121.1	6.0	119.7	6.2	114.0	6.3

**Table 8. Statistical results of fixed effect parameters: yields (2008-2010).<sup>[a]</sup>**

Production Characteristic	2008			2009			2010		
	Moisture Treatment (MT)	Irrigation Type (IT)	MT × IT	Moisture Treatment (MT)	Irrigation Type (IT)	MT × IT	Moisture Treatment (MT)	Irrigation Type (IT)	MT × IT
Total yield	*	*	NS	NS	NS	NS	NS	NS	NS
Marketable yield	*	*	NS	NS	NS	NS	NS	NS	NS
Reds	*	*	NS	NS	NS	NS	*	NS	NS

<sup>[a]</sup> NS = not significant; asterisk (\*) indicates significance at  $p \leq 0.05$ .



**Figure 1. Irrigation depth vs. fruit yield for 2008 to 2010.  $Y$  = yield and  $x$  = irrigation depth.**

review of the data revealed that two of the eight 70% FC plots had yields of 130.7 and 176.2 Mg ha<sup>-1</sup>. The 176.2 Mg ha<sup>-1</sup> plot was the highest yielding of the 32 plots and received the highest equivalent depth of irrigation (249.2 mm). These two exceptionally high yields therefore skewed the average yield for the 70% FC treatment, making it the treatment with the highest yield.

The 60% FC treatment represented a 74% depletion of the AWC. The plant physiological stress associated with the 60% FC treatment undoubtedly contributed to the lowest yields. The 80% FC treatment, which corresponded to 36% depletion in AWC, had unusually low yields. Five of the eight 80% FC plots had yields of less than 100 Mg ha<sup>-1</sup> and resulted in an average yield that was lower than for the 70% FC treatment.

In relation to irrigation type, all the surface drip irrigated treatments produced higher yields than the corresponding buried drip irrigated treatments in 2008. This difference was attributed to the depth of the buried drip lines. During the 2008 season, the buried drip lines were installed at 20 cm below the surface; however, in 2009 and 2010, they were installed at 15 cm below the surface. Processing

tomato has an effective rooting depth of approximately 30 to 40 cm in loamy sand. It was believed that the drip line placement at 20 cm depth may have limited the wetting pattern and the capillary rise during irrigation, limiting the amount of water available to the crop within the rooting zone, particularly between the 0 to 10 cm depths. Visual observations indicated that the soil surface for the buried drip irrigated plots were often dry, which may have been due to low capillary rise. However, the surface irrigated plots provided a longer period for the irrigation water to move through the root zone, and by extension supplied a greater amount of water at the effective rooting depth of 30 cm. This undoubtedly contributed to the significant difference between the yields of the irrigation types. Tan et al. (2003) reported similar trends but attributed the higher yields for surface irrigated plots to root intrusion into the subsurface emitters, preventing uniform water distribution. Phene et al. (1987) reported conflicting results for surface and buried (45 cm below the surface) drip irrigated processing tomatoes grown in clay loam soils in California. Manual and machine harvest yields for subsurface drip treatments were 10.3% and 17% greater, respectively, than for high-frequency surface drip treatments, and 29.2% and 24.1% greater than for low-frequency surface drip treatments. However, it must be noted that both the soil type and the depth of the drip lines were different in the current experiment.

#### FRUIT YIELDS IN 2009 AND 2010

In 2009, the average fruit yield for the four treatments ranged from 101.6 to 105 Mg ha<sup>-1</sup>. The highest and lowest average yields were from the -30 kPa and 91% FC treatments, respectively. There was very little variation in yields between treatments during the 2009 season; therefore, no statistical significance was obtained among the moisture treatments, irrigation types, or the interaction between the two. This was primarily due to the fact that the soil moisture depletion levels of the four treatments were

reduced, such that three of the four treatments were within the MAD of 30% to 40% AWC. The fourth moisture treatment (74% FC) was just 5% outside the range of the MAD recommended by Doorenbos and Pruitt (1977) and had approximately the same average yield as the 91% FC treatment. This may be due to the masking effect of the effective rainfall (186 mm), which would have minimized the effect of the 74% FC treatment particularly. The 91% FC treatment (depletion of 15% AWC) also had a lower yield than the -30 kPa treatment (depletion of 22% AWC). It is possible that the 91% FC treatment created too wet a soil environment in the rooting zone, thus negatively impacting plant growth and development, resulting in a slightly lower plant yield. Figure 1 indicates that the maximum yield was reached with a seasonal irrigation depth of approximately 150 mm. After this point, fruit yield decrease with increase irrigation.

In 2010, the average fruit yield ranged from 104.4 to 121.1 Mg ha<sup>-1</sup>. The two treatments that fell within the MAD range of 30% to 40% AWC had higher yields, while the two treatments that fell outside the MAD range had lower yields. The 2010 season produced the highest average yields among the treatments (113.7 Mg ha<sup>-1</sup>). It is strongly believed that the large effective rainfall depth over the growing season (May to August, with each month having greater rainfall than its 30-year average, and the seasonal average being 43% greater than the 30-year average) masked the effects of the treatments and thus reduced the plant stress, particularly during the critical plant growth stages. It is therefore not surprising that there were no statistical differences between moisture treatments, irrigation types, or the interaction between them.

The results over the three years indicated a direct relationship between irrigation volume and yield (fig. 1). Total and marketable production increased with increasing irrigation depth. Similar results were reported by Machado and Oliveira (2005), Machado et al. (2000), and Sezen et al. (2010) for both surface and buried drip irrigation. Machado and Oliveira (2005) obtained comparable yields, ranging from 78.8 to 141.7 Mg ha<sup>-1</sup> and from 69.9 to 130.1 Mg ha<sup>-1</sup> for total and marketable yields, respectively. Warner et al. (2007) reported total and marketable yields of 130 to 173.3 Mg ha<sup>-1</sup> and 126.7 to 168.5 Mg ha<sup>-1</sup>, respectively. Zhang et al. (2010) reported total and marketable yields of 64 to 166.7 Mg ha<sup>-1</sup> and 56 to 138 Mg ha<sup>-1</sup>, respectively. In

each of the three years, the treatments with ≤40% AWC depletion generally produced higher yields, thus validating the recommendation of Doorenbos and Pruitt (1977). The -30 kPa treatment represented depletion in the AWC ranging from 22% to 24% VWC over the three years. Apart from the anomaly in 2008, when the 70% FC treatment had the highest yield, the tension treatment had the highest yields in 2009 and 2010, despite not being statistically significant. The threshold values (-10 and -30 kPa) for the fill and refill levels were comparable to other studies done with processing tomatoes. Marouelli and Silva (2007) used soil water thresholds (SWT) ranging from 5 to 120 kPa and found that the best yields were obtained when irrigation was performed at SWT thresholds of 35, 12, and 15 kPa during the vegetative, fruit development, and maturation growth stages, respectively, at Embrapa Vegetables, in Brasília, Brazil. Hartz and Hanson (2005) recommended thresholds in the range of 20 to 35 kPa up to fruit maturation and a range of 40 to 50 kPa after this point. However, these values were for deep clayey soils in California. It is also worth noting that while these researchers varied the threshold values over the developmental stages of the crop, the current experiment kept the threshold values constant throughout the growing season.

## FRUIT QUALITY AND IWUE

### FRUIT QUALITY IN 2008

Fruit quality parameters and their statistics are summarized in tables 9 and 10. In 2008, the average fruit weights from the different treatments reflected a trend similar to that of yields and had similar statistical results. Two of the seven fruit qualities parameters (weight and soluble solids) indicated statistical significance among the treatments and were influenced by the irrigation scheduling program and the irrigation depths for the various treatments over the season. All other fruit quality factors indicated no statistical significance. In relation to irrigation type, there was statistical significance for fruit weight, firmness, soluble solids, and Brix yields, which may be attributed to the fact that the surface irrigated plot received more irrigation than the buried irrigated plots. There was no statistical significance for the interaction between moisture treatment and irrigation type for all seven fruit quality parameters.

**Table 9. Fruit quality parameters per treatment from 2008 to 2010 (standard errors shown in parentheses).**

Year	Treatment	Weight (g)	Firmness (MPa)	Area (cm <sup>2</sup> )	Color (Agron)	Soluble Solids (°Brix)	pH	Brix Yield (Mg ha <sup>-1</sup> )	IWUE (kg m <sup>-3</sup> )
2008	60% FC	39.7 (1.2)	2.31 (0.04)	16.1 (0.8)	18.7 (0.5)	6.0 (0.2)	4.2 (0.01)	5.38 (0.3)	30.57 (2.2)
	70% FC	43.9 (2.0)	2.25 (0.06)	16.5 (0.4)	18.3 (0.5)	5.2 (0.2)	4.3 (0.01)	5.96 (0.2)	34.50 (1.8)
	80% FC	40.3 (1.4)	2.29 (0.05)	15.2 (0.5)	18.1 (0.8)	6.0 (0.2)	4.2 (0.01)	5.74 (0.6)	33.13 (2.8)
	-30 kPa	45.4 (2.0)	2.23 (0.06)	16.6 (0.4)	17.7 (0.5)	5.3 (0.2)	4.3 (0.01)	5.96 (0.1)	32.15 (1.2)
2009	74% FC	-	2.03 (0.05)	14.1 (0.4)	19.6 (0.4)	4.8 (0.1)	4.3 (0.01)	4.81 (0.1)	37.56 (1.9)
	82% FC	-	1.88 (0.06)	14.8 (0.1)	19.7 (0.3)	4.7 (0.1)	4.3 (0.01)	4.77 (0.2)	31.77 (4.2)
	91% FC	-	1.85 (0.05)	14.7 (0.3)	19.9 (0.5)	4.5 (0.1)	4.3 (0.01)	4.44 (0.2)	28.10 (2.9)
	-30 kPa	-	1.89 (0.04)	15.0 (0.3)	19.9 (0.4)	4.6 (0.1)	4.3 (0.01)	4.76 (0.2)	30.85 (1.4)
2010	55% FC	45.1 (1.9)	2.68 (0.09)	15.9 (0.4)	18.8 (0.6)	5.2 (0.2)	4.2 (0.01)	5.22 (0.2)	28.90 (1.1)
	70% FC	50.7 (1.7)	2.59 (0.12)	17.3 (0.3)	18.6 (0.4)	4.9 (0.2)	4.3 (0.01)	5.15 (0.3)	27.13 (1.9)
	85% FC	54.1 (1.6)	2.73 (0.11)	18.0 (0.2)	18.3 (0.4)	4.5 (0.1)	4.3 (0.01)	5.22 (0.2)	24.55 (1.2)
	-30 kPa	56.3 (2.4)	2.74 (0.08)	18.6 (0.5)	18.0 (0.4)	4.5 (0.1)	4.3 (0.00)	5.38 (0.3)	24.41 (1.2)

**Table 10. Statistical results of fixed effect parameters: Fruit quality (2008-2010).**<sup>[a]</sup>

Production Characteristic	2008			2009			2010		
	Moisture Treatment (MT)	Irrigation Type (IT)	MT × IT	Moisture Treatment (MT)	Irrigation Type (IT)	MT × IT	Moisture Treatment (MT)	Irrigation Type (IT)	MT × IT
Fruit weight	*	*	NS	NA	NA	NA	*	NS	NS
Fruit firmness	NS	*	NS	NS	NS	NS	NS	NS	NS
Fruit area	NS	NS	NS	NS	NS	NS	*	NS	NS
Fruit color	NS	NS	NS	NS	NS	NS	*	NS	NS
Soluble solids	*	*	NS	NS	NS	NS	*	NS	NS
Fruit pH	NS	NS	NS	NS	NS	NS	NS	NS	NS
Brix yield	NS	*	NS	NS	NS	NS	NS	NS	NS
IWUE	NS	NS	NS	NS	NS	NS	NS	NS	NS

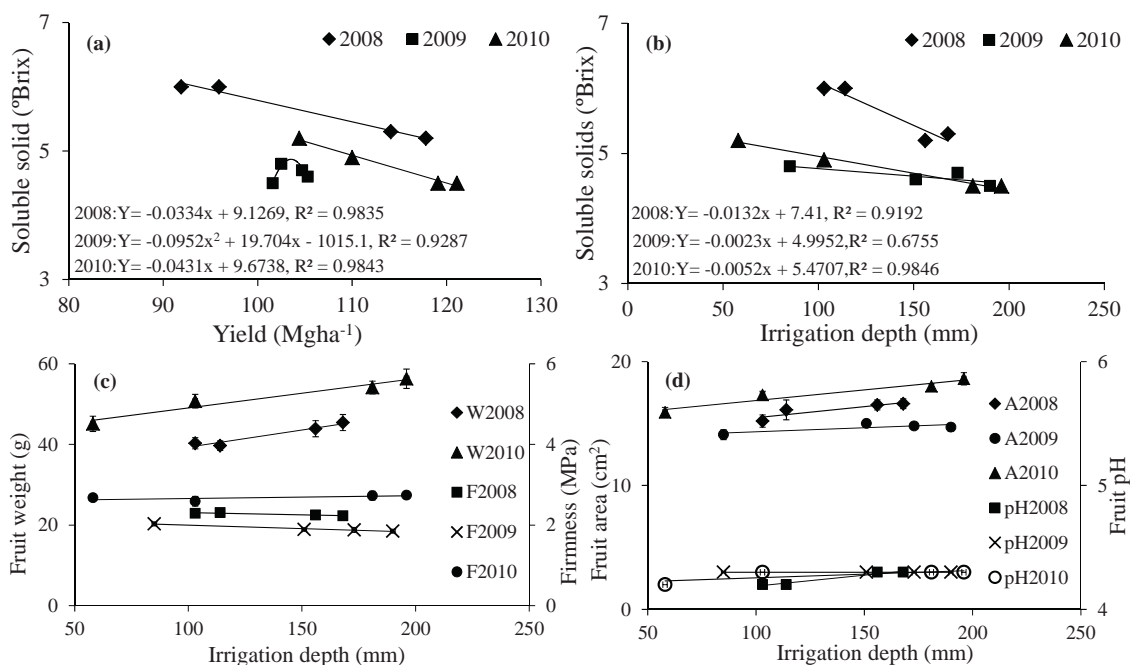
<sup>[a]</sup> NA = not available; NS = not significant; and asterisk (\*) indicates significance at  $p \leq 0.05$ .

### FRUIT QUALITY IN 2009 AND 2010

In 2009, individual fruit weight was not measured due to unavailability of equipment. All six fruit quality parameters indicated no statistical significance among the moisture treatments, irrigation types, or the interaction between the moisture treatment and irrigation type. Three factors contributed to the absence of statistical significance in 2009. First, the 2009 moisture treatments reflected a less stressed moisture irrigation scheduling program than the 2008 and 2010 treatments. The range between FC treatments was very small, and hence the differences were minimal. Three of the four moisture treatments were within the 30% to 40% MAD range, and the other (74% FC) represented a depletion in the AWC of 45%. The second factor was the depth at which the buried drip lines were installed in 2009. The drip lines were raised by 5 cm, to a depth of 15 cm below the soil surface, which facilitated better moisture distribution and improved yields for the buried drip irrigated plots as compared to 2008. The third factor was the relatively effective rainfall over the growing season, which masked the treatment effects, particularly for the 74% FC treatment.

In 2010, the range between the FC treatments was increased; as a result, greater differences were realized in the fruit quality parameters among the moisture treatments. Four of the seven fruit quality parameter (weight, size, color, and soluble solids) indicated statistical significance between the moisture treatments. There was no statistical significance among the irrigation types and the interaction of moisture treatment and irrigation type. As was the case in 2009, the shallower depth of the drip lines accounted for the improved yields for the buried irrigated plots, compared to the surface drip irrigated plots, which resulted in the absence of statistical significance among irrigation types.

Over the three years, both irrigation depth and fruit yield had converse effects on soluble solids (figs. 2a and 2b). Fruit weight and fruit size showed positive relationships with irrigation depth, while fruit firmness showed an inverse relationship with irrigation depth. Fruit pH showed no real relationship with moisture content (figs. 2c and 2d). In all three years, there was a negative relationship between irrigation depth and soluble solids. A negative correlation was also obtained between fruit yield and soluble solids in 2008 and 2010. A similar negative correlation was reported



**Figure 2. (a) Yield vs. soluble solids, (b) irrigation depth vs. soluble solids, (c) irrigation depth vs. fruit weight (W) and firmness (F), and (d) irrigation depth vs. fruit area (A) and pH. In 2a and 2b,  $Y$  = soluble solids and  $x$  = yield and irrigation depth for 2a and 2b, respectively. In 2c and 2d, error bars are standard errors of means.**



by Machado and Oliveira (2005). In 2009, the soluble solids increase with increasing irrigation to a critical irrigation depth (150 mm) and subsequently decreased with increasing irrigation. It was found that pH was not statistically affected by irrigation depth, type, or treatment interaction. This was consistent with observations by Machado et al. (2003) and Davis et al. (1985). The Brix yield, which is the product of fruit soluble solids and the marketable yield, ranged from 4.77 to 5.96 Mg ha<sup>-1</sup> over the three years. Statistical significance was realized only between irrigation types for 2008. This was attributed to the large difference in yield between the irrigation types. Machado and Oliveira (2005) reported soluble solids, pH values, and Brix yields ranging from 4.51 to 6.07 °Brix, from 4.30 to 4.38, and from 4.20 to 5.85 Mg ha<sup>-1</sup>, respectively, which were comparable with the ranges reported in this research.

A possible water conservation approach to large-scale processing tomato production in Canada is to remunerate tomato growers based on soluble solids content (SSC). High SSC values are highly desirable for processing, and processors pay a premium for tomatoes with a high SSC (Dumas et al., 1994; Iddo, 2008). A practical approach would be to establish a threshold Brix yield, which can be achieved by setting a slightly higher fruit SSC. This would necessitate a reduction in irrigation water application, which would inevitably result in a slight decrease in yield. However, the net Brix yield would be the same. The savings to the grower would be in reduced irrigation cost. The unused surplus water could be used for further expansion. However, further research work in this area is necessary in a Canadian context.

The IWUE ranged from 30.5 to 35 kg m<sup>-3</sup>, from 28.10 to 37.6 kg m<sup>-3</sup>, and from 24.4 to 28.9 kg m<sup>-3</sup> for 2008, 2009, and 2010, respectively. In 2008, the 70% FC treatment yielded the highest IWUE (table 9). This was due to the unusually high yield for the 70% FC treatment. The 2008 IWUE showed a positive relationship with irrigation depth (fig. 3). Both of these factors can be attributed to the unusually higher yields for the 70% FC treatment. In 2009 and 2010, the lowest moisture treatment produced the highest IWUE due to the comparatively lower irrigation depth. Both of these years showed a converse relationship with irrigation depths, such that an increase in irrigation

depths resulted in a corresponding decrease in IWUE. Although the effective rainfall was substantial over the three years, the IWUE was not significantly affected by moisture treatment, irrigation type, or their interaction. This trend is in agreement with Machado and Oliveira (2005), who reported IWUE values ranging from 20.2 to 22.8 kg m<sup>-3</sup>. These values are substantially lower than that obtained in this research and are attributed to the higher water applied (irrigation and rainfall), which ranged from 326.2 to 644.0 mm.

## CONCLUSIONS

In southwestern Ontario, irrigated agriculture is a prerequisite for large-scale field tomato production. Commercial yield was higher for the moisture treatments in which the greatest quantity of water was applied. This also corresponded to the moisture treatments in which the depletion level in soil moisture was ≤40% AWC. The tension-based treatment (-30 kPa) produced the highest yields in two of the three years of the experiment. The 2008 results indicated that surface drip irrigation produced significantly higher yields than buried drip irrigation for each of the moisture treatments, which was attributed to the depth of the buried drip lines in 2008. However, in the other two years (2009 and 2010), there was no statistically significant difference in yield between the surface drip and buried drip irrigated plots after the buried drip lines were raised from 20 to 15 cm depth. The rainfall distribution, particularly during the critical growth periods of the crop, may have contributed by masking the moisture treatments and reducing crop stress. The fruit quality parameters of greatest interest were weight, size, firmness, soluble solids content, and Brix yield. The heavier and larger fruits were associated with the wetter moisture treatments. Brix yield showed statistical significance only between the surface and buried irrigation systems in 2008. The IWUE showed no statistical significance between the moisture treatments, irrigation types, or their interaction for each of the three years; however, 2010 had the lowest IWUE, which was due primarily to higher rainfall than in the previous two years. As a water conservation approach for large-scale processing tomato production in Canada, it is worth exploring the possibility of remunerating tomato growers based on soluble solids content (SSC) and Brix yield instead of on total yields.

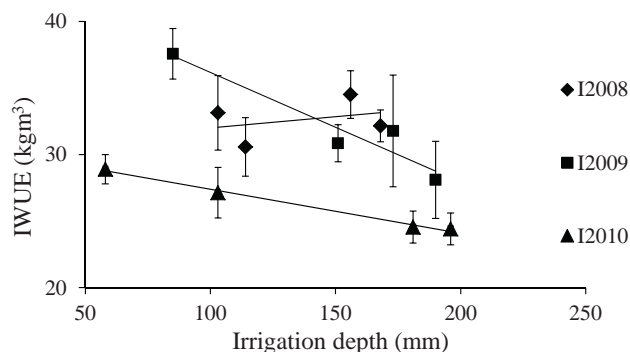


Figure 3. Irrigation depth vs. IWUE for 2008 to 2010. Error bars are standard errors of the means.

## ACKNOWLEDGEMENTS

This research has been made possible by a grant from the Research and Innovation Branch of the Ontario Ministry of Agriculture, Food, and Rural Affairs. Appreciation is extended to Mr. Wayne Palichuck for the use of his farm and facilities. Special thanks to Apurva Gollamudi for his assistance in setting up the field instrumentation, data collection, and editing the manuscript. Thanks are also extended to the technical staff at McGill University.



## REFERENCES

- AAFC. 2006. AAFC science and innovation strategy. Ottawa, Ontario, Canada: Agriculture and Agri-Food Canada. Available at: [www4.agr.gc.ca/AAFC-AAC/display-afficher.do?id=1175602657035#s4](http://www4.agr.gc.ca/AAFC-AAC/display-afficher.do?id=1175602657035#s4). Accessed 18 September 2012.
- Allen, R. G., L. S. Pereira, D. Raes, and M. Smith. 1998. Crop evapotranspiration: Guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper No. 56. Rome, Italy: Food and Agriculture Organization of the United Nations.
- Campbell, G. S., and M. D. Campbell. 1982. Irrigation scheduling using soil moisture measurements: Theory and practice. *Adv. Irrig. Sci.* 1: 25-42.
- Davis, K. R., C. J. Phene, R. L. McCormick, R. B. Hutmacher, and D. W. Meek. 1985. Trickle frequency and installation depth effects on tomatoes. In *Proc. Third Intl. Drip/Trickle Irrigation Congress*, 896-901. St. Joseph, Mich.: ASAE.
- Doorenbos, J., and W. O. Pruitt. 1977. Guidelines for predicting crop water requirements. FAO Irrigation and Drainage Paper No. 24. Rome, Italy: Food and Agriculture Organization of the United Nations.
- Dumas, Y., C. Leoni, Portas, C. A. M., and B. Bièche. 1994. Influence of water and nitrogen availability on yield and quality of processing tomato in the European Union countries. *Acta Hort.* 376: 185-192.
- Fares, A., and A. K. Alva. 2000. Soil water components based on capacitance probes in a sandy soil. *SSSA J.* 64(1): 311-318.
- Fares, A., H. Hamdhani, V. Polyakou, A. Dogan, and H. Valenzuela. 2006. Real-time soil water monitoring for optimum water management. *J. American Water Resources Assoc.* 42(6): 1527-1535.
- Fereres, E., D. A. Goldhamer, and L. R. Parsons. 2003. Irrigation water management of horticultural crops. *HortScience* 38(5): 1036-1042.
- Haise, H. R., and R. M. Hagan. 1967. Soil, plant, and evaporation measurements as criteria for scheduling irrigation. In *Irrigation of Agricultural Land*, 577-604. Agronomy Monograph No. 11. R. Hagan, H. Haise, and T. Edminster, eds. Madison, Wisc.: ASA.
- Hanson, B. R., R. B. Hutmacher, and D. M. May. 2006. Drip irrigation of tomato and cotton under shallow saline ground water conditions. *Irrig. Drain. Syst.* 20(2-3): 155-175.
- Hartz, T., and B. Hanson. 2005. Drip irrigation and fertigation management of processing tomato. Davis, Cal.: University of California, Vegetable Research and Information Center.
- Hartz, T. K., P. R. Johnstone, D. M. Francis, and E. M. Miyao. 2005. Processing tomato yield and fruit quality improved with potassium fertigation. *HortScience* 40(6): 1862-1867.
- Howell, T. A. 2001. Enhancing water use efficiency in irrigated agriculture. *Agron. J.* 93(2): 281-289.
- Hudson, B. D. 1994. Soil organic matter and available water capacity. *J. Soil and Water Cons.* 49(2): 189-194.
- Iddo, K. 2008. Yield quality and irrigation with saline water under environmental limitations: The case of processing tomatoes in California. *Agric. Econ.* 38(1): 57-66.
- Johnstone, P. R., T. K. Hartz, M. LeStrange, J. J. Nunez, and E. M. Miyao. 2005. Managing fruit soluble solids with late-season deficit irrigation in drip-irrigated processing tomato production. *HortScience* 40(6): 1857-1861.
- Jones, B. J., Jr. 2007. *Tomato Plant Culture: In the Field, Greenhouse, and Home Garden*. Boca Raton, Fla.: CRC Press.
- Keeney, D. R., and D. W. Nelson. 1982. Part 2: Nitrogen: Inorganic forms. In *Methods of Soil Analysis*, 643-698. Agronomy Monograph No. 9. A. L. Page, ed. Madison, Wisc.: ASA and SSSA.
- LeBoeuf, J., C. Tan, and V. A. 2007. Irrigation scheduling for tomatoes: An introduction. OMAFRA Factsheet. Guelph, Ontario, Canada: Ontario Ministry of Agriculture, Food, and Rural Affairs.
- Liu, K., T. Q. Zhang, and C. S. Tan. 2011. Processing tomato phosphorus utilization and post-harvest soil profile phosphorus as affected by phosphorus and potassium additions and drip irrigation. *Canadian J. Soil Sci.* 91(3): 417-425.
- Lukangu, G., M. J. Savage, and M. A. Johnston. 1999. Use of sub-hourly soil water content measured with a frequency-domain reflectometer to schedule irrigation of cabbages. *Irrig. Sci.* 19(1): 7-13.
- Machado, R. M. A., and M. D. R. G. Oliveira. 2005. Tomato root distribution, yield, and fruit quality under different subsurface drip irrigation regimes and depths. *Irrig. Sci.* 24(1): 15-24.
- Machado, R. M. A., L. R. A. Olivera, and C. A. M. Portas. 2000. Effect of drip irrigation and fertilizer on tomato rooting patterns. *Acta Hort.* 537: 313-320.
- Machado, R. M. A., M. do Rosario, G. Oliveira, and C. A. M. Portas. 2003. Tomato root distribution, yield, and fruit quality under subsurface drip irrigation. *Plant and Soil* 255(1): 333-341.
- Marouelli, W., and W. Silva. 2007. Water tension thresholds for processing tomatoes under drip irrigation in central Brazil. *Irrig. Sci.* 25(4): 411-418.
- OHCRSC. 2006. Ontario horticultural crop research and service committee report. Guelph, Ontario, Canada: Ontario Ministry of Agriculture, Food, and Rural Affairs.
- Olsen, S. R., C. V. Cole, F. S. Watannabe, and L. A. Dean. 1954. Estimation of available phosphorus by extraction with sodium bicarbonate. Circular No. 939. Washington, D.C.: USDA.
- Patanè, C., and S. L. Cosentino. 2010. Effects of soil water deficit on yield and quality of processing tomato under a Mediterranean climate. *Agric. Water Mgmt.* 97(1): 131-138.
- Peters, D. B. 1965. Part 1: Physical and mineralogical properties: Water availability. In *Methods of Soil Analysis*, 279-285. Agronomy Monograph No. 9. C. A. Black, ed. Madison, Wisc.: ASA.
- Phene, C., K. Davis, R. Hutmacher, and R. McCormick. 1987. Advantages of subsurface irrigation for processing tomatoes. *Acta Hort.* 200: 101-114.
- Ratliff, L. F., J. T. Ritchie, and D. K. Cassel. 1983. Field-measured limits of soil water availability as related to laboratory-measured properties. *SSSA J.* 47(4): 770-775.
- SAS. 2007. The SAS system for Windows. Release 9. 0. Cary, N.C.: SAS Institute, Inc.
- Schwab, G. O., D. D. Fangmeier, J. M. Elliot, and K. R. Frevert. 1993. *Soil and Water Conservation Engineering*. New York, N.Y.: Wiley.
- Sentek. 2003. Access tube installation guide: Version 1.0 for EnvironSCAN, EnviroSMART, and Diviner 2000. Stepney, South Australia: Sentek Sensor Technologies.
- Sezen, S. M., G. Celikel, A. Yazar, S. Tekin, and B. Kapur. 2010. Effect of irrigation management on yield and quality of tomatoes grown in different soilless media in a glasshouse. *Sci. Res. and Essay* 5(1): 041-048.
- Shock, C. C., and F.-X. Wang. 2011. Soil water tension, a powerful measurement for productivity and stewardship. *HortScience* 46(2): 178-185.
- Shock, C. C., E. B. G. Feibert, L. D. Saunders, and E. P. Eldredge. 2001. Automation of subsurface drip irrigation for crop research. In *Proc. World Congress of Computers in Agriculture and Natural Resources*, 809-816. St. Joseph, Mich.: ASABE.
- Shock, C. C., A. Pereira, B. Hanson, and M. Cahn. 2007. Vegetable irrigation. In *Irrigation of Agricultural Crops*, 535-606.

- Agronomy Monograph No. 30. R. Lescano and R. Sojka, eds. Madison, Wisc.: AASA, CSSA, and SSSA.
- Stanley, C. D., and D. N. Maynard. 1990. Vegetables. In *Irrigation of Agriculture Land*, 921-950. Agronomy Monograph No. 30. B. A. Stewart and D. R. Nielsen, eds. Madison, Wisc.: ASA.
- Statistics Canada. 2011. Fruit and vegetable production: February 2011. Catalog No 22-003-X. Ottawa, Ontario, Canada: Statistic Canada.
- Tan, C. S., and W. D. Reynolds. 2003. Impacts of recent climate trends on agriculture in southwestern Ontario. *Water Resources J.* 28(1): 87-97.
- Tan, C. S., T. Q. Zhang, W. D. Reynolds, C. F. Drury, and A. Liptay. 2003. Farm-scale processing tomato production using surface and subsurface drip irrigation and fertigation. ASABE Paper No. 032092. St. Joseph, Mich.: ASABE.
- Thompson, R. B., M. Gallardo, L. C. Valdez, and M. D. Fernandez. 2007a. Using plant water status to define threshold values for irrigation management of vegetable crops using soil moisture sensors. *Agric. Water Mgmt.* 88(1-3): 147-158.
- Thompson, R. B., M. Gallardo, L. C. Valdez, and M. D. Fernández. 2007b. Determination of lower limits for irrigation management using *in situ* assessments of apparent crop water uptake made with volumetric soil water content sensors. *Agric. Water Mgmt.* 92(1-2): 13-28.
- Warner, J., C. S. Tan, and T. Q. Zhang. 2007. Water management strategies to enhance fruit solids and yield of drip-irrigated processing tomato. *Canadian J. Plant Sci.* 87(2): 345-353.
- Zhang, T. Q., C. S. Tan, K. Liu, C. F. Drury, A. P. Papadopoulos, and J. Warner. 2010. Yield and economic assessments of fertilizer nitrogen and phosphorus for processing tomato with drip fertigation. *Agron. J.* 102(2): 774-780.

*Tools for the implementation of Integrated Water Resources Management (IWRM) in the Caribbean: the legacy of the Caribbean Water Initiative (CARIWIN)*

**Catherine Senecal and Chandra A. Madramootoo**

McGill University, Brace Centre for Water Resources Management, Macdonald Campus of McGill University, 21111 Lakeshore Road, Ste-Anne-de-Bellevue, QC, Canada, H9X 3V9, [catherine.senecal@mcgill.ca](mailto:catherine.senecal@mcgill.ca)

While many countries and regional authorities in the Caribbean have embraced the concept of Integrated Water Resources Management (IWRM) and recognized its guiding principles as beneficial, few possessed the capacity to implement it since its enunciation in the Dublin Principles of 1992. The Caribbean Water Initiative (CARIWIN) has endeavoured over a six-year period, since 2006, to build capacity in a collaborative process with national governments, regional and international agencies. Working in conjunction with the Caribbean Institute for Meteorology and Hydrology (CIMH), CARIWIN produced three tools to support the implementation of the key components of IWRM: the National Water Information Systems, the Caribbean Drought and Precipitation Monitoring Network, and Community Water Strategies.

## **1. Introduction**

*“Scarcity and misuse of fresh water pose a serious and growing threat to sustainable development and protection of the environment. Human health and welfare, food security, industrial development and the ecosystems on which they depend, are all at risk, unless water and land resources are managed more effectively in the present decade and beyond than they have been in the past.”*

The Dublin Statement on Water and Sustainable Development,  
January 31, 1992, Dublin, Ireland

The International Conference on Water and the Environment, which issued the Dublin Statement on Water and Sustainable Development, was pivotal in the global paradigm shift towards Integrated Water Resources Management, now known by its widely applied acronym IWRM. At that juncture in history, there was a formal call for fundamental new approaches to the assessment, development and management of freshwater resources (UN, 1992). Four guiding principles were proposed and are summarized as follows: effective water management requires a holistic approach; a participatory approach; the recognition of women’s vital role; and the recognition of the economic value of water. The benefits of implementing IWRM were detailed as including alleviation of poverty and disease; protection against natural disasters (drought and flood); water conservation and reuse; sustainable urban development; food security and improved rural water supply; protection of aquatic ecosystems; and resolving water conflicts (UN, 1992).

With 10 years of reflection on the matter, when reunited at the 2002 Johannesburg World Summit on Sustainable Development, nations of the world, including Caribbean states, agreed to have IWRM plans in place by 2005 (Cashman, 2012). While Cashman concludes in his 2012 review of regional progress that there is an acceptance among professionals and decision-makers of a real need for re-orientation of water sector governance towards IWRM, he concedes that the

reality in the Caribbean is that the process is long and fraught with difficulties. Despite this delay in policy development and institutional reform, Caribbean Water Initiative (CARIWIN) has supported efforts in the region to expand the knowledge base with respect to water resources and to strengthen capacity of institutions and personnel in order to facilitate IWRM implementation. The legacy of CARIWIN is explored in its five areas of focus: (i) decision-support tools, (ii) professional development, (iii) partnerships, (iv) research, and (v) dissemination of knowledge.

## **2. CARIWIN Overview**

The Caribbean Water Initiative is a collaborative project, designed to address the complex challenges of water management in the Caribbean region. CARIWIN was granted \$1,000,000.00 Canadian Dollars over six years (2006-2012) by the Canadian International Development Agency, through its University Partnerships in Cooperation and Development Program. The main institutional partners are the Brace Centre for Water Resources Management of McGill University and the Caribbean Institute for Meteorology and Hydrology (CIMH), with collaborating national agencies in Grenada (Ministry of Agriculture), Guyana (Ministry of Agriculture), and Jamaica (Water Resources Authority). These collaborating national agencies coordinated the involvement of other government agencies, non-governmental organizations, and local communities in CARIWIN sponsored events. Through CIMH, CARIWIN piloted capacity building initiatives in IWRM, aimed at the national, local government and community levels in Jamaica, Guyana and Grenada.

In order to achieve the goal of IWRM in the Caribbean, CARIWIN proposed the strengthening of the CIMH, a regional organization whose mandate is to provide training and capacity development in climatology and water management to CARICOM member states. By integrating the IWRM approach into CIMH training and capacity development initiatives, the project will have a significant multiplier effect throughout the Caribbean. The main outputs, the decision-support tools, have achieved a high level of acceptance and have already propagated beyond the three pilot countries. Notably, these decision-support tools were not ready-made external solutions being imported to the region, but rather they were developed and nurtured with Caribbean stakeholders using an inclusive and collaborative approach. CARIWIN has the potential to make a long-term contribution to poverty reduction, gender equality, improved governance and local ownership, better health through safe water provision and enhanced environmental sustainability through more effective water management.

## **3. Decision-support Tools**

In order to make well-informed decisions, managers and policy-makers require timely information which is relevant to the needs of society and based on sound data. Data and information to support decision-making are vital to effective water management (IHP, 2000). CARIWIN endeavoured to meet these needs in the region through the development of three decision-support tools: the National Water Information Systems, the Caribbean Drought and Precipitation Monitoring Network, and Community Water Strategies.

### ***3.1 National Water Information Systems***

A National Water Information System (NWIS) is a centralized database for a country's water-related information and a powerful decision-support tool. The NWIS is currently operational in

five Caribbean countries, with several others in the initial phases of development or investigation. The system is preferred for its ease of adaptability and expandability, in addition to the fact that it is based on open source software. The NWIS has the potential to facilitate sharing of data and information for decision support at national (cross-sectoral) and regional levels.

Through a collaborative process including stakeholders from varied sectors and representing both data-users groups and data-collectors groups, the Government of Grenada developed and implemented its National Water Information System (NWIS) early in 2009. The Grenada NWIS has now become an official repository for all hydrologic, climate, land, watershed, infrastructure and water-related data in the country (St-Jacques *et al.*, 2010). Contrary to the traditionally sector-segregated data storage and analysis, the NWIS is a tool able to support decision-making in line with the guiding principle of adopting a holistic approach to water management. The development of the Grenada NWIS took into consideration the recommendations which emanated from the United Nations Food and Agricultural Organization's prior experience in the region. One noteworthy recommendation in this context was that the types of data collected should be expanded beyond the technical, to include socio-economic, legal, environmental and institutional so that all stakeholders could have ready access to information to assist in decision making in all aspects of the water sector (Fletcher-Paul *et al.*, 2008). Water governance has been strengthened in Grenada as water managers, now armed with the NWIS, are better able to assess the resource; share information with other national agencies and stakeholders; clearly report pertinent information; and positively influence policy (St-Jacques *et al.*, 2010).

### ***3.2 Caribbean Drought and Precipitation Monitoring Network***

The role of the CIMH as a regional provider of meteorological and hydrological services has been strengthened by the Institute taking the lead in drought monitoring for the Caribbean Basin. The drought monitor, posted on the CIMH website for easy access to information by all stakeholders, is a valuable decision-support tool for water resources management. The information products posted by the monitor are generated through the analysis of raw meteorological data provided by member countries. These are used to produce a visual indication of wet or dry conditions based on the Standard Precipitation Index (SPI) which is accompanied by a verbal description of the conditions. Currently, SPIs are calculated for 1,3,6 and 12 month time intervals. This Caribbean Drought and Precipitation Monitoring Network (CDPMN) was launched in 2009 and provides a Drought Alert service for the region, which has already served to inform water managers and other socio-economic sectors of the severity of the dry spell that had gripped the region in October 2009 until the end of May 2010.

Being able to anticipate drought through constant monitoring via indices and other climatic indicators as well as biological observations allows for mobilization of resources and preparation that increases the coping capacity of communities and nations (Trotman *et al.*, 2008b). The lesson of the 2009-2010 drought was that on the Caribbean basin scale, rainfall indices gave reasonable notice as to the existence and severity of the drought, however it is clear that national scale monitoring incorporating other indicators (agricultural and hydrological) is also needed (Trotman *et al.*, 2011). One national monitor has been launched in the region, for Barbados, and efforts are being made to develop others. The CDPMN affords an opportunity for a participatory process, between CIMH, national and local governments and communities, to propose adaptation strategies and new Community Water Strategies which consider the extremes of drought and flood for water resource management. Although CARIWIN did not pursue activities at the



community level, CIMH will carry on with these efforts within other programs such as the European Union funded Caribbean Agro-meteorological Initiative that focuses on farming communities.

### ***3.3 Community Water Strategies***

CARIWIN produced a documented framework to guide the development and implementation of Community Water Strategies (CWS) in the Caribbean, based on IWRM principles. The focus for the preparation of the framework was on synthesizing key components of IWRM at the community level, and lessons learned from case studies, into a manageable process to guide the development of CWS. The framework is presented as a four-phase process involving assessment, planning, implementation and monitoring, where each phase is broken down into several sub-components and associated with specific expected outputs (McGill, 2009a). This was also expressed and refined within country-specific documents for each of the CARIWIN partner countries, i.e. Grenada, Guyana and Jamaica. The CARIWIN CWS Background Documents for Grenada, Guyana and Jamaica each examine water resources management at the national level for the country in question, establishing the context for the CWS in the pilot watershed. The information is summarized from country policy documents, strategies, and on-going programs and projects (McGill, 2009b).

These four reference documents, the framework plus the three country-specific documents, were used as the basis for discussion during the CARIWIN Regional Seminar held in Guyana, January 14-15, 2010. At this event, a capacity-building exercise conducted with national partners focussed on the prioritization of the pilot community needs; preliminary identification of key players; and the steps to implementation for each country to move forward with the CWS. Each partner country was thus trained in formulating a CWS for their pilot community and CARIWIN transferred the onus onto the individual countries to lead any further development.

## **4. Professional Development**

Continuous updating of the knowledge base among water sector personnel is critical to maintaining a skilled work force with appropriate abilities and know-how to tackle the problems at hand. CARIWIN professional development activities have each specifically targeted a particular group of beneficiaries ranging from technical personnel, senior water managers, policy-makers, to CIMH personnel and other stakeholders such as researchers, and personnel from international or regional agencies. In all, there were 301 individual professional development opportunities offered through 23 events held in seven countries. In addition to the 301 opportunities, an additional 136 persons were present in the role of presenters or experts. Details of the professional development offered through CARIWIN are listed in Annex 1. While many of these events were of the workshop, seminar, training session, or short course type, three were official programs offered by post-secondary educational institutions.

### ***4.1 Academic programs***

With additional financing from the University of Guyana, CARIWIN assisted in the development of a new academic program in the form of a Post-Graduate Certificate in Water Resource Management offered through the University of Guyana. The program consists of three courses: Hydrology and Water Resources Management; Drainage and Irrigation; and Hydraulic

Structures and Geotechnical Investigation. The Stabroek News (2012) reported the Minister of Agriculture's justification for the new program as a need for renewed efforts in equipping the ministry's human resource with the necessary skills to build capacity.

With additional financing from McGill University, CARIWIN assisted in the development of the on-line professional development program titled Integrated and Adaptive Water Resources Management and Governance offered through McGill University. This program takes a holistic look at water resources management at the watershed and sub-watershed level with a specific focus on watershed planning techniques, adaptive management strategies, capacity building approaches, water governance, and water business risk assessment to strengthen integrated water resources management capacities globally. The method of web-based course delivery facilitates the participation of working professionals around the globe.

#### ***4.2 Human Resources and Technical Capacity Building***

The other events (training, workshop, seminar, high-level meetings etc.) had corollary benefits other than conferring knowledge such as: networking opportunities, exposure to the latest advancements, and consensus-building for development of decision-support tools. In addition, several of these events served as models which CIMH has adopted for use and replicated within other regional initiatives.

Practical and hands-on learning was an integral part of the professional development activities delivered. A review of documentation indicated that the Caribbean faces many challenges in managing its water resources in a socially acceptable, environmentally sustainable, and economically efficient manner due to problems which include *inter alia* poor hydrological data collection and analysis (IHP, 2000). CARIWIN installed pilot hydrometric stations and weather stations in each partner country, i.e. Grenada, Guyana and Jamaica. These were used for demonstration and training purposes with respect to installation, operation, and maintenance of the equipment as well as data management and the integration of the data into Water Information Systems for storage and analysis. These in-country stations allowed CIMH to train their member country personnel on site, and therefore in greater numbers than would normally have been possible at the CIMH campus in Barbados. In addition to the use of these stations for training purposes, the stations are generating additional data uploaded to the national networks for hydrological and meteorological monitoring. To similar ends, CARIWIN also provided hand-held water quality monitoring equipment to CIMH and to each partner country.

### **5. Partnerships**

In setting recommendations for the implementation of Integrated Water Resources Management in the region, twenty-two Caribbean countries participated in a workshop organized by the Caribbean Council for Science and Technology, with support of the Inter-American Development Bank. One out of the four main recommendations of the workshop focused on the need to develop strategic partnerships and networks for fostering information sharing and exchange (IDB, 1999). Furthermore, in the preamble to the St. George's Declaration of Principles for Environmental Sustainability in the OECS, the signatories recognize that their situation requires a collaborative approach:

*“ACCEPTING that their small size and limited technical, financial and human resources constrain the ability of Member States to unilaterally undertake all the actions required to achieve their own goals and their international obligations regarding sustainable development...” (OECS,2006).*

In this regard, the development and propagation of the National Water Information Systems included strategic partnership building between the national governments and their counterparts in other Caribbean countries and the CIMH, FAO, CEHI, GEF-IWCAM, GWI and UWI. In the case of the Caribbean Drought and Precipitation Monitor, the collaboration was extended to include CDEMA, NOAA, Agriculture and Agri-Food Canada, EU and the National Drought Mitigation Center. These partnerships were regional agencies were extremely significant in that they sparked synergies and led to coordinated efforts. For example, the FAO contributed financial support, technical backstopping, and knowledgeable professional personnel towards the development of the NWIS in Grenada, which could not have materialized in the same form or within the same time frame without. Another example is the continual collaboration with CEHI throughout the development of the NWIS and beyond as CEHI took on the role of champion for a coordinated regional approach to water management by leading the organization of a Caribbean Water Information System Experts Meeting in 2011 with representatives from twelve countries in attendance. This leveraging of technical, financial and human resources from institutions with compatible objectives made it possible to achieve results with an impact of far greater magnitude than would belie the modest initial investment.

## **6. Research**

Scientific research conducted through CARIWIN contributed to the advancement of effective water resources management in the Caribbean region and the conferment of five Master of Science degrees and one Doctoral degree. The results advance the knowledge base in the areas of drought monitoring and household water supply and storage.

### **6.1 Drought monitoring**

Trotman and Mehdi (2008) concluded that the determination of drought indices was an important step in the development of an integrated water resources management program. Research has since evaluated the application of established drought indices such as the Standard Precipitation Index (SPI) and Normalized Difference Vegetation Index (NDVI) to the Caribbean context, particularly, the applicability of both the NDVI and SPI for representing soil water conditions at three sites in Jamaica was evaluated by Richards and Madramootoo. The SPI was found to have good correlation and therefore to have significant potential as a drought indicator (Richards *et al.*, 2010). Furthermore, Stoute, Trotman and Charlerly studied the application of internationally accepted drought indices PDSI and PSI to the practice of drought prediction in Barbados. The calculated indices showed a high correlation with soil water content analyses. With these research results, the CIMH adopted the use of the SPI as the drought indicator of choice for the Caribbean Drought and Precipitation Monitor published on the CIMH website.

The applicability of the Soil Water Assessment Tool (SWAT) was evaluated for use in an agricultural watershed in Jamaica. Despite the fact that the model had some difficulties in simulating high-runoff events, it was determined that SWAT is a suitable model for use in simulating streamflow and holds much potential for future agricultural water resources planning

in all Jamaica watersheds, in particular, as it can be used to determine agricultural water-saving strategies (Richards and Madramootoo, 2010). Richards and Madramootoo (2010) also published the land use and soil parameters that were used for this model so that they may be used as a reference in the development of future hydrologic simulations in Jamaica.

Research in the area of drought monitoring is contributing to the region's effort to move from a response-driven approach towards a strategic approach focused on prevention and mitigation of drought disasters.

### ***6.2 Household water supply and storage***

Research was conducted in an aboriginal community in rural Guyana in order to suggest best practices and provide information based on IWRM principles which may serve to inform decision-making for water supply and water safety investments. A study to compare the rates of adoption at the household level between three drinking water treatment options was conducted using the Biosand slow-sand-filter, the ceramic candle filter, and the application of a chlorine product marketed as Chlorosol. Of the three treatments tested, it was found that ceramic candle filters performed best in terms of the rate of adoption (Young-Rojanschi *et al.*, 2009).

An investigation was also conducted in the same community on the feasibility of scaling-up Domestic Rain Water Harvesting (DRWH) to improve water security. Intven argued that DRWH systems were shown to be a relatively low cost option for universally improving a households' geographical and temporal access to a water source. As domestic rainwater harvesting (DRWH) was already accepted and used by the majority of community residents, albeit mainly in an informal manner, it was concluded that facilitating the installation of formal DRWH systems will have a large impact on household water security.

Both of these studies at the household level incorporated the guiding principles of a) adopting a participatory approach to formulating recommendations for water management solutions and of b) the recognition of women's vital role in the provision of water. This research contributes to the body of knowledge in the region with regards to the results, but also with regards to the process of applying IWRM principles within research projects in order to inform water management decisions.

## **7. Dissemination of Knowledge**

Knowledge gained, but not shared is of little use to the advancement of society. The Caribbean Water Initiative has, to date, published 10 papers and posters, made 8 conference presentations in the Caribbean and around the world, published innumerable freely accessible resources on the CARIWIN website as well as provided new information on the CIMH website. This is part of an informal capacity building strategy encompassing public awareness which seeks to disseminate experiences and learning related to IWRM. One example in which CARIWIN has emphasized public awareness raising is with respect to the economic value of water. Trotman in particular clarifies the link between drought disasters and their resulting impacts on various economic sectors. Specifically, he highlighted that the impacts of the 2009-2010 drought in the Caribbean included water rationing; major crop losses; increases in food prices; drop in hydro power production; increased water delivery costs for irrigation; *etc.* (Trotman *et al.*, 2011). Trotman also reminded that during the drought of 1998, Guyana experienced water rationing, cessation of

logging and river transport in some places and the loss of livestock (Trotman *et al.*, 2009). Further, he informed that, due to drought and major losses in the sugar sector, the Jamaican government provided the sector a US\$100 million assistance package in 1997. This recognition of the economic value of water is incorporated into dissemination of knowledge activities. As a result, the region has developed and implemented a drought monitor, the CDPMN, in order to assist planning and improve resilience in the region.

## 8. Conclusions

The Caribbean region, despite the uneven completion of institutional and policy reforms required to allow IWRM practices to flourish, has experienced recent advancements towards the IWRM paradigm. In particular, the recognition of increased effectiveness in the management of water resources was experienced in various spheres of the Caribbean Water Initiative. Several examples were provided to demonstrate the improvements to water management where the four guiding principles were embraced *i.e.* of adopting a holistic approach; adopting a participatory approach; recognizing women's vital role in water management; and recognizing the economic value of water. Of particular promise for the region are the decision-support tools such as the National Water Information Systems, the Caribbean Drought and Precipitation Monitoring Network, and the Community Water Strategy frameworks which are all three shown to be meeting real and immediate needs of the region, but which will undoubtedly need to be continuously revisited and upgraded in order to maintain relevance in the water management context. CARIWIN supported and nourished these decision-support tools for IWRM by promoting professional development opportunities, partnership building, scientific research, and the dissemination of knowledge. IWRM is a valid and necessary objective for protection against natural disasters (drought and flood); food security and improved rural water supply; and poverty alleviation among other positive outcomes and it is in the best interest of Caribbean nations to strive towards the strengthening of IWRM decision-support tools.

## References

Cashman, A. 2012. Water Policy development and governance in the Caribbean: an overview of regional progress. *Water Policy*. 14:14-30.

Fletcher-Paul, L.M., C.A. Madramootoo, and H. Thomas. 2008. A review of Water Information Systems in the English-speaking Caribbean - challenges and lessons learnt. Caribbean Environmental Forum. 24-25 June, 2008. St. Georges, Grenada.

Inter-American Development Bank (IDB). 1999. Integrated Water Resources Management: Institutional and Policy Reform. IDB, Washington.

International Hydrological Programme. 2000. An overview of selected policy documents on water resources management that contributed to the design of HELP (Hydrology for the Environment, Life and policy). Technical Documents in Hydrology, No. 38. UNESCO, Paris.

McGill University (McGill). 2009a. Community Water Strategies: A framework for implementation.

[https://secureweb.mcgill.ca/cariwin/sites/mcgill.ca.cariwin/files/cws\\_framework\\_v2.pdf](https://secureweb.mcgill.ca/cariwin/sites/mcgill.ca.cariwin/files/cws_framework_v2.pdf)



McGill University (McGill). 2009b. CARIWIN Community Water Strategy Background Document: Great River Watershed, Grenada.  
[https://secureweb.mcgill.ca/cariwin/sites/mcgill.ca.cariwin/files/cws\\_grenada\\_december2009.pdf](https://secureweb.mcgill.ca/cariwin/sites/mcgill.ca.cariwin/files/cws_grenada_december2009.pdf)

Organisation of Eastern Caribbean States (OECS). 2006. St. George's Declaration of Principles for Environmental Sustainability in the OECS. pp 30. OECS, St. Lucia.

Richards, J. and C.A. Madramootoo. 2010. An investigation into the feasibility of using SWAT at the sub-basin level for simulating hydrologic conditions in Jamaica. Caribbean Environmental Forum. 21-25 June, 2010. Montego Bay, Jamaica.

Richards, J., C.A. Madramootoo and A. Trotman. 2010. The development of the SPI and NDVI for 3 study sites in Jamaica, with an investigation into their use in understanding soil water during water stressed conditions in Jamaica. Caribbean Environmental Forum. 21-25 June, 2010. Montego Bay, Jamaica.

Stabroek News 2010. Twenty-four engineers for hydrology studies under UG-McGill pact. February 1, 2010. Georgetown, Guyana.  
<http://www.stabroeknews.com/2010/archives/02/01/twenty-four-engineers-for-hydrology-studies-under-ug-mcgill-pact/>

St-Jacques, M-C., C. Senecal, T. Thompson, A. Haiduk, A. Trotman, and C.A. Madramootoo. 2010. National Water Information Systems: A tool to support Integrated Water Resources Management in the Caribbean. Caribbean Environmental Forum. 21-25 June, 2010. Montego Bay, Jamaica.

Trotman, A., D. Farrell, and C. Cox. 2011. Drought early warning and risk reduction: A case study of the Caribbean drought of 2009-2012. World Climate Research program Drought Interest Group Meeting. 2-4 March, 2011. Barcelona, Spain.

Trotman, A., B. Mehdi, A. Gollamudi, C. Senecal. 2008a. Drought and Precipitation Monitoring for Enhanced Integrated Water Resources Management in the Caribbean. Caribbean Environmental Forum. 24-25 June, 2008. St. Georges, Grenada.

Trotman, A., L. Pologne, S. Stoute, B. Mehdi, C. Senecal, and A. Gollamudi. 2008b. A proposed approach to monitoring and assessing drought in the Caribbean. Second Turkey-Israel Workshop on Drought Monitoring and Mitigation. June 16-29, 2008. Turkey.

Trotman, A., A. Moore, S. Stoute. 2009. The Caribbean Drought and Precipitation Monitoring Network: the concept and its progress, in Climate Sense. Tudor Rose, UK.

Young-Rojanschi, C., C.A. Madramootoo, L. Intven, and C. Senecal. 2009. Comparing three HWTS options in St. Cuthbert's Mission, Guyana. International Research Colloquium of the Network to Promote Household Water Treatment and Safe Storage. September 21-23, 2009, Ireland.

United Nations (UN). 1992. The Dublin Statement on Water and Sustainable Development. International Conference on Water and the Environment. January 31, 1992. Dublin Ireland.

Professional Development Activities Offered Through CARIWIN							
year	type	duration	event title	location	target	topics	reach
2007	Symposium	24h 0m 0s	CARIWIN Symposium	Barbados	senior water managers	CARIWIN; experiences and challenges of IWRM in the Caribbean; FAO's initiatives on water management in the Caribbean; integrated watershed and coastal areas management activities by CEHI	30
2007	Workshop	72h 0m 0s	Senior Administrators IWRM Workshop	Barbados	senior water managers	principles of IWRM; incorporation of IWRM into national decision-making bodies and strategies; use of data for value-added products; community involvement in IWRM	12
2007	Training	48h 0m 0s	Gender Equality: Training Session and Strategy Development	Canada	CIMH	importance and challenges of gender equality in water management; particularities of GE in the Caribbean: design of a GE strategy for CARIWIN; role of women in drinking water supply and hygiene	1
2007	Training	336h 0m 0s	IWRM Training	Canada	CIMH	technical aspects of IWRM	2
2007	Short course	336h 0m 0s	Advanced IWRM Course	Barbados	senior water managers	hydrological data; water quality; GIS; flood analysis; climate change; drought management; GE in water management.	15
2007	Short course	336h 0m 0s	Hydrometeorology and Water Quality Field Course	Guyana	technical personnel and senior water managers	water quality monitoring fundamentals and equipment; hydrometeorology measurements; case studies in water quality; database management	25
2008	Training	336h 0m 0s	IWRM Training	Canada	CIMH	IWRM; hydrological data; water quality; soil moisture monitoring; modelling tools; climate change; drought management.	2
2008	Short course	120h 0m 0s	Hydrometric Equipment Installation and Training	Grenada	technical personnel	water level measurements, reference marks, cross-sections, flow measurements, rating curves, levels and levelling, uses of hydrological data, Practical exercises in calculation of river flows.	15
2008	Workshop	48h 0m 0s	Grenada NWIS Workshop	Grenada	stakeholders	data management; precipitation and drought monitoring; national water information systems	19
2008	Short course	336h 0m 0s	Hydrometric Equipment Installation and Training	Guyana	technical personnel	river level measurements using staff gauges and water level recorders; relationship between river level and river flow and the rating curve; Installation of hydrometric equipment; practical exercises in river flow measurement and cross-section surveys.	11
2008	CIMH program	8 months	Mid-level Technician, Class III	Barbados	technical personnel	instrumentation; observation; statistics; GIS	1

Professional Development Activities Offered Through CARIWIN						
year	type	duration	event title	location	target	reach
2009	Workshop	48h 0m 0s	2nd Senior Administrators Workshop	Barbados	senior water managers	18
2009	Training	72h 0m 0s	NWIS Launch and Training Session	Grenada	technical personnel and senior water managers	21
2009	Training	48h 0m 0s	Hydrometric Training	Grenada	technical personnel	21
2009	Training	336h 0m 0s	IWRM Training	Canada	CIMH	2
2009	Training	72h 0m 0s	Hydromet equipment training	Guyana	technical personnel	2
2009	Seminar	48h 0m 0s	CARIWIN Regional Seminar	Guyana	senior water managers and other stakeholders	24
2010	Workshop	168h 0m 0s	SWAT: Computer modelling workshop	Canada	CIMH	1
2010	Post-graduate program	3 months	University of Guyana Post Graduate Certificate in Water Resource Management	Guyana	senior water managers	20
2011	Training	72h 0m 0s	NWIS Training in Grenada	Grenada	senior water managers	10
2011	Workshop	96h 0m 0s	Water Information Systems Experts Meeting	St. Lucia	senior water managers and other stakeholders	25
2012	Workshop	72h 0m 0s	Drought Training Workshop	Jamaica	technical personnel and senior water managers	20
2012	Post-graduate program	3 months	Integrated and Adaptive Water Resources Planning, Management and Governance	on-line	senior water managers	4





**International Journal of Applied Research and Technology**  
ISSN 2277-0585

Publication details, including instructions for authors and subscription information:  
<http://www.esxpublishers.com>

**Physical and Processing Characteristics of Some Popular Rice Varieties in Nigeria**

Danbaba, N.<sup>1</sup>, Nkama, I.<sup>2</sup>, Manful, J.<sup>3</sup>, Michael, N.<sup>4</sup>, Amaka, M. O.<sup>4</sup> and Badau, M. H.<sup>2</sup>

<sup>1</sup>National Cereals Research Institute, Badeggi, Niger State, Nigeria

<sup>2</sup>University of Maiduguri, Nigeria

<sup>3</sup>Africa Rice Center (AfricaRice), Cotonou, Benin

<sup>4</sup>McGill University, Quebec, Canada, H9X 3V9

Available online: January 31, 2013

To cite this article:

Danbaba, N., Nkama, I., Manful, J., Michael, N., Amaka, M. O. and Badau, M. H. (2013). Physical and Processing Characteristics of Some Popular Rice Varieties in Nigeria. *International Journal of Applied Research and Technology*. 2(1): 64 – 73.

**PLEASE SCROLL DOWN FOR ARTICLE**

*This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan, sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden*

*The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instruction, formulae and analysis should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.*



## Physical and Processing Characteristics of Some Popular Rice Varieties in Nigeria

Danbaba, N.<sup>1</sup>, Nkama, I.<sup>2</sup>, Manful, J.<sup>3</sup>, Michael, N.<sup>4</sup>, Amaka, M. O.<sup>4</sup> and Badau, M. H.<sup>2</sup>

<sup>1</sup>National Cereals Research Institute, Badeggi, Niger State, Nigeria

<sup>2</sup>University of Maiduguri, Nigeria

<sup>3</sup>Africa Rice Center (AfricaRice), Cotonou, Benin

<sup>4</sup>McGill University, Quebec, Canada, H9X 3V9

(Received: 04 January 2013 / Accepted: 21 January 2013 / Published: 31 January 2013)

### Abstract

The objective of this work was to determine the physical and processing qualities of nine (9) popular rice varieties cultivated in Nigeria as a strategy for developing sustainable rice processing technology. Physical properties and milling attributes including paddy length (L), width (W), thickness (T), Mass ( $W_{1000}$ ), surface area ( $S_a$ ), aspect ratio ( $A_r$ ), sphericity ( $S_c$ ) and volume (V) and percentage head rice, broken grains, milling degree and milling recovery were measured at moisture content of 12.86% dry weight basis and the following results were obtained: the average L, W, L/W ratio, T and  $W_{1000}$  were 9.46mm, 2.57mm, 3.70, 2.25mm and 31.77g respectively, while calculated physical properties, GMD, V,  $S_c$ ,  $A_r$  and  $S_a$  were 3.78mm, 28.89cm<sup>3</sup>, 40.05%, 26.67% and 44.86cm<sup>2</sup> respectively. The mean milling characteristics significantly ( $p=0.05$ ) varied in terms of hull, head rice recovery, percentage broken fraction, milling degree and milling recovery as 19.18%, 69.97%, 15.45%, 2.65% and 79.06% respectively. These characteristics are primarily considered when developing rice processing machinery and processes, while hierarchical analyses indicate close similarity between the varieties in terms of their milling yield.

**Keywords:** Paddy, Physical Properties, Rice Milling, Head Rice Recovery, Nigeria

Corresponding author:

E-mail: [zirbabs@yahoo.com](mailto:zirbabs@yahoo.com)

ISSN 2277-0585

© 2013 Exxon Publishers. All rights reserved.

## Introduction

Rice (*Oryza sativa* L.) is one of the most important cereals cultivated world-wide, constituting the basic food for large number of human beings, sustaining two-third of the world population (Zhout *et al.*, 2002; Varnamkhasti *et al.*, 2007) and the most important crop in the world in terms of total developing world production ( $480 \times 10^6$  tons of rough rice) and number of consumers (2.50 billion) dependent on it as their staple food (Juliano, 2007). According to FAOSTAT (2011), the world's rice production increased from 520 million tons in 1990 to over 678 million tons in 2009, while the Nigeria's rice production increased from 3.2 million metric tons in 1995 to about 4.6 million metric tons in 2009.

The Nigerian government in 2010 reviews its policy and strategies for sustainable rice production and mitigating factors responsible for poor quality of locally processed rice, marketing, distribution and value addition in an integrated manner (NRDS, 2010). The increasing importance of rice for sustainable food security in Nigeria and complexity of modern processing technology, demands comprehensive information on the quality attributes of popular rice varieties cultivated in the country. The quality of rough rice delivered to rice mills plays a critical role in determining the quality of the final milled rice (SAED, 2000; Kergna and Coulibaly, 2001). To obtain better quality milled rice, modeling dynamic abrasion in milling operations and for designing suitable polishing and cleaning systems, the knowledge of physical properties of diverse rice produced in a given locality is important (Mohapatra and Bal, 2004). For this purpose that Varnamkhasti *et al.*, (2007) suggested the evaluation of grain dimension (length, width, length/ratio), volume, surface area, thousand grain weight, aspect ratio, sphericity, porosity and angle of repose of rough rice as quality of prime importance. These properties, Varnamkhasti *et al.*, (2007) and Suthar and Das (1996) reported are principal attributes influencing the design and evaluation of rice processing operations including drying, husking, whitening and polishing as well as grading machines, storage and grain moving equipment.

Physical and processing properties of rough rice grown in different parts of the world have been previously studied (Morita and Singh, 1979; Steffe and Singh, 1980; Reddy and Chakraverty, 2004; Correa *et al.*, 2007; Varnamkhasti, *et al.*, 2007; Ghadge and Prasad, 2012). But little or no data exist in scientific literatures on the physical and processing characteristic of common rice varieties cultivated in Nigeria. The current study therefore was designed and conducted to determine some of the paddy physical and processing properties such as linear dimensions, paddy weight, surface area, volume, geometric mean diameter, sphericity, aspect ratio, percentage hull and milling performance of some popular rice varieties grown in Nigeria as a strategy to generate scientific data for the improvement of technical system rice processing operation for the general enhancement of the competitiveness of locally milled rice.

Rudimentary handling methods and technologies used for rice in Nigerian and indeed Sub Saharan African countries result in significant qualitative and quantitative losses. Mechanization of the rice postharvest system has been suggested using intermediate technologies. But, if the machines and operations are not properly designed in line with the attributes of the product intended to be processed, low quality product may still be produced. Physical properties of food materials like paddy grains is of fundamental importance in designing of harvesting, transporting, and dimensioning of storage facilities and procedures and operations of different post-harvest equipment (Ghadge, *et al.*, 2008; Ghadge and Prasad, 2012). Marketing of agricultural products after harvesting depend on its physical qualities, for instance, 1000-weight of rough rice grain have been used for estimating the head rice yield of a given rice variety and this attribute (percentage of whole grain) is one of the most important parameter considered by processors, marketers and consumers in the rice value chain. The present study therefore, provide some important information on the dimensional and processing characteristics of most common rice varieties grown in Nigeria, including local varieties, previously improved varieties and recently released rice varieties as a strategy for improving rice post-harvest value-chain and competitiveness of locally processed rice in the country. Such information is intended also to provide improvement in communication on local and international rice trade.

## Materials and Methods

The paddy rice (9 rice varieties including 3 earlier released varieties (FARO 44, FARO 52 and FARO 57), 4 popular farmers varieties (*Jamila*, *Jeep*, *Yardas* and *Kwandala*) and 2 (NERICA L-19 (FARO 60) and NERICA L-34 (FARO 61) newly released varieties) used in this study was obtained from the Rice Breeding Laboratory of the National Cereals Research Institute (NCRI), Badeggi and Seed Stores of Kano State Agricultural and Rural Development Agency (KNARDA), Kano, Nigeria. They are from the 2010 harvest (December), they were harvested by panicle and threshed and cleaned manually and packaged in a polyethylene bags. The paddy when received from the sources were cleaned manually and packaged in a paper bag at room temperature ( $30 \pm 2^\circ\text{C}$ ) until required for analysis, while the analysis was carried out at the Rice Grain Quality Laboratory of NCRI. The initial moisture contents of the paddy were determined using the hot air oven method.

Method described by Danbaba *et al* (2011) was adopted for parboiling of the rice samples with slight modifications. 1kg each of cleaned samples were measured into a small cloth bag and labeled before placing into a boiling water in a heated vessel provided with lid and the samples stirred such that the final water temperature was around  $75 \pm 2^\circ\text{C}$  and left over night to soak. After 18 hrs, the water was drained and steeped paddy steamed in a laboratory parboiling vessel (locally fabricated at NCRI). Steaming was monitored till when the paddy hull start to open and the process halted. Steamed rice were thereafter sprayed out on the laboratory bench to dry ( $32 \pm 2^\circ\text{C}$ ) to moisture content of about 14-16%. About 250g dried parboiled rice samples were dehulled with a laboratory rice huller (Satake Engineering Co. Ltd., Japan) and the resulting brown rice milled with McGill No. 2 laboratory mill and the head rice yield obtained after grading using

Satake testing rice grader (TRG 05A) using suitable indented cylinder to remove broken kernel of size shorter than three-fourths of the whole grain length. The milling quality attributes were then evaluated using the formulas below (IRRI, 1996).

$$\text{Head Rice Recovery (HRR)\%} = \frac{\text{Weight of paddy grains}}{\text{Weight of paddy sample}} \times 100 \quad (1)$$

$$\text{Percentage Broken (PBR)\%} = \frac{\text{Weight of broken grains}}{\text{Weight of paddy sample}} \times 100 \quad (2)$$

$$\text{Percentage milling degree (PMD)\%} = \frac{\text{Weight of milled rice}}{\text{Weight of broken rice}} \times 100 \quad (3)$$

$$\text{Percentage Milling recovery (PMR)\%} = \frac{\text{Weight of milled rice}}{\text{Weight of sample used}} \times 100 \quad (4)$$

$$\text{Percentage hull (PH)\%} = \frac{\text{Weight of hull removed}}{\text{Weight of paddy sample}} \times 100 \quad (5)$$

Rough rice grain were randomly selected (100 paddy per sample) and their three principal dimensions, length (L), width (W) and thickness (T) were measured using digital venire caliper (Mitutoyo Corporation, Japan) to an accuracy of 0.01mm. The geometric mean diameter ( $D_g$ ) was calculated from the principal dimensional figures evaluated above using the relationship given by Mohsenin (1986) as:

$$D_g = (LWT)^{1/3} \quad (6)$$

Thousand weight of paddy ( $W_t$ ) was determined by counting 100 rice seed and weighing them in an electronic balance and the readings multiplied by 10 to give the mass of 1000-seeds (Ghasemi *et al.*, 2007). The sphericity, which defines the characteristics shape of the grain relative to that of sphere of the same volume, was determined as described by Mohsenin (1986) as:

$$Sc = \frac{(LWT)^{1/3}}{L} \times 100 \quad (7)$$

The aspect ratio ( $A_r$ ) used in the classification of grain shape (Varnamkhasti *et al.* 2007) was calculated as outlined by Maduako and Faborode (1990) as:

$$Ar = \frac{W}{L} \times 100 \quad (8)$$

The surface area ( $S_a$ ) of rough rice were calculated using the relationship in equation (9) given by Varnamkhasti *et al.* (2007):

$$Sa = \frac{\pi BL^2}{(2L-B)} \quad (9)$$

Where:

$$A = \sqrt{WT} \quad (10)$$

While the equation (11) stated by Jain and Bal (1997) was adopted for the evaluation of paddy volume (V):

$$V = 0.25 \left[ \left( \frac{\pi}{6} \right) L(W + T)^2 \right] \quad (11)$$

Data generated from the above analysis were expressed as means, standard deviation and coefficient of deviation of 100 determinations for paddy length, width, thickness, geometric mean diameter, surface area and aspect ratio, while 10 readings were reported for volume and 1000-paddy weight. For analyzing the obtained data, Microsoft Excel Software (Microsoft Cooperation, USA) was used. While milling quality data were subjected to analysis of variance (ANOVA) using CROPSTAT 7.2.3 (International Rice Research Institute, Philippines) and hierarchical cluster analyses was performed using SPSS 16.0.

## Results and Discussions

Summary of the all the physical parameters measured and determined are presented in Tables 1 to 3. The means, standard deviations, frequency distribution in terms of skewness and kurtosis and coefficient of variation for local rice varieties are

presented in Table 1, the moisture content of the different rice varieties at the time of this experiment was  $12.86 \pm 0.78\%$  dry weight basis. This moisture content can help to suggest the stability in storage of the paddy rice.

The longitudinal dimensions or length (L) for the different paddy rice ranged from 9.13 to 9.60 mm with a mean value of 9.37 mm. For the width (W), the distribution is between 2.50 and 2.71 mm with a mean value of 2.63 mm. However, coefficient of variation (CV) for length and width are between 4.58 to 6.575 and 4.53 to 7.75% respectively. Though Jeep has the lowest CV for length (4.58%), it has the highest (7.75%) in terms of width (Table 1). The mean thickness ranged from 2.01 (*Jamila*) to 2.38 mm (*Kwandala*) with CV ranging from 6.30 to 9.95% in *Kwandal* and *Jamila* respectively. Kramer (1951) measured these characteristics from 35 rice varieties (rough and brown). He established length-thickness and width-thickness ratios. The CV of the measurements was used to indicate the variability of dimension or uniformity of grain within variety. Postharvest gravimetric properties of paddy rice such as length, width, thickness and surface has also been reported as characteristics of agricultural products that are important in solving problem involving the design of specific machines for processing and handling (Prashant and Prasad, 2012; Varnamkhasti, *et al.*, 2007; Arora (1991); Arana (2007); Kachru, *et al.*, (1994). What shape is to be assumed for the material and which dimension is to be employed in calculations are two of the first technical problems to be solved before solving problems related to separation of agricultural products (Danbaba *et al.*, 2012). The result of the analysis of the principal diameters of the paddy shows frequency distribution in terms of kurtosis and skewness among the length, width and thickness as -0.84, -0.54 and 0.72 and -0.11, 0.38 and -0.90 respectively. Skewness characterizes the degree of symmetry of a distribution around its means and therefore, positive skewness indicates a distribution with an asymmetric tail extending towards more values, while negative skewness indicates a distribution with an asymmetric tail extending towards more negative values. Zero values indicate symmetrical distribution (Simonyan, *et al.* 2007). Kurtosis on the other hand characterizes the relative peakness or flatness of a distribution compared to the normal distribution. Positive kurtosis indicates leptokurtic distribution. Negative kurtosis indicates platykurtic distribution. Zero values indicate normal or mesokurtic distribution.

The geometric mean diameter of the axial dimensions is usual in the estimation of the projected area of a particle moving in the turbulent or near- turbulent region of an air stream. This projected area of the particle (paddy) is generally indicative of its pattern of behavior in a flowing fluid such as air, as well as the ease of separating extraneous materials from the material lot during cleaning by pneumatic means (Omobuwajo *et al.*, 1999; Varnamkhasti, *et al.*, 2007). In this study, geometric mean diameter was between 3.57 and 3.88mm among local rice varieties and 3.69 to 3.86mm and 3.77 and 3.79mm in earlier improved varieties and the NERICA rice varieties. In all the 3 classes of paddy reported in this study, the mean geometric diameter are uniform (3.77mm), indicating that they may have similar behavior in a flowing fluids.

The weight of agricultural products is exploited in the design of cleaning equipment using aerodynamic force (Oje and Ugbor, 1991). In rice, it is an important quality attribute considered in trade at the international market. 1000-paddy weight in this study varied among local varieties from 27.22 to 38.00g with a mean value of 31.23g (Table 1), 26.60 to 37.80g and 32.61 and 36.60g among improved varieties (Table 2 and 3). The one thousand paddy mass is a useful index in measuring the relative amount of milling outturn, dockage or foreign material in a given lot of paddy, and the amount of shriveled or immature kernels (Luh, 1980). This result is higher than 19.47 to 21.57g reported by Varnamkhasti, *et al.* (2007) in *Sazandegi*, a popular variety in Iran. The frequency distributions of the various paddy physical characteristics are shown in Table 1, 2 and 3 in terms of skewness and kurtosis. Skewness shows lack of symmetry about the mean in a frequency distribution of the measured and calculated physical properties, while kurtosis shows the extent the frequency distribution is concentrated about the mean. Positive skewness means the frequency distribution is skewed to the right. Kurtosis is the degree of peakedness of the distribution with respect to a normal distribution which is mesokurtic. High kurtosis means the frequency is concentrated about the mean.

The sphericity of different rice varieties ranges from 37.77 to 42.22% among the local varieties and 37.43 to 39.34% in earlier released improved varieties, while newly introduced varieties has sphericity value of 39.24 and 44.60%. While the aspect ratio varied from 26.45 to 29.20% among the local varieties (Table 1) and 20.23 to 24.98% in earlier released improved varieties and 26.47 to 31.29% in the newly introduced NERICA rice varieties. Ghadge and Prasad (2012) reported sphericity value of 33.33 to 43.57% with a mean value of 26.58% in a popular Indian rice variety, PR-106. While Varnamkhasti *et al.* (2007) reported sphericity of *Sazandegi* rice variety of Iran to range between 37.30 to 43.20% with a mean value of 39.88%. Wolfe and Tatepo (1972) in a study of the terminal velocity of chopped forage materials reported that sphericity of an agricultural material like rice paddy is always below unity and that the more regular an object is, the lower the sphericity. Oje (1993) also suggested that the higher the sphericity, the higher the tendency of the material to slide than roll during processing. In the present study, the less 50% sphericity results of all the tested rice samples is an indication of the tendency of the paddy to slide during processing. The aspect ratios in all the samples varies from 26.45 to 29.20% (Table 1), 20.23 to 24.98% (Table 2) and 26.47 and 31.29% (Table 3) in local, earlier introduced improved and newly introduced improved varieties respectively. This result is in agreement with reports of Varnamkhasti *et al.* (2007). Low aspect ratio, Eke *et al.*, (2007) reported is a tendency of seeds to being oblong in shape. Therefore, with aspect ratio of less than or equal to 30%, paddy rice is more likely not to roll but slide during conveying or other postharvest operations. Sphericity and aspect ratio data Eke *et al.*, (2007) reported are critical for the design of hoppers for rice mills, separators in destoners and conveyors in medium and large scale rice processing mills. The low CVs of these parameters in all the samples confirms that they are were uniformly dispersed about their mean values. Surface area ranged from 40.22 to 46.96 cm<sup>3</sup> with a mean value of 44.98cm<sup>3</sup> (Table 1), 42.82 to 47.72cm<sup>3</sup> with mean of 44.56cm<sup>3</sup> (Table 2) and 44.83 and 45.34cm<sup>3</sup> (mean 45.09) (Table 3) in local, earlier introduced and newly improved varieties respectively. Surface area in

paddy rice determines the shape and this attributes Alonge and Adigun (1999) reported is an indication of the way the paddy will behave on oscillating surface during processing.

Milling quality is composed of several factors that directly affect the value of rough rice. It encompasses the total amount of milled rice recovered after milling (total milled-rice yield) and the total amount of whole kernels recovered after milling (head-rice or whole rice yield) (Bergman, et al., 2004). During rice milling, paddy is first thoroughly cleaned and then the husk removed using existing huskers. The performance of a rice husker is not only governed by the working parameters of a husker (engineering properties) but also the physical and morphological characteristics (varietal properties) of the paddy to be processed (Saeid et al, 2007). Grains to be subjected to husking and other milling processes must be first meet some physical attribute quality that will meet the requirement of the processes (Kunze 1985; Ancheta and Aandales, 1990; Saeid, 2007). Milling characteristics of the rice varieties evaluated in this study in terms of head rice recovery (HRR), percentage broken rice, percentage milling recovery and percentage of hull are presented in Table 4. Milling turnout of all the rice varieties was satisfactory, ranging from 74.27 to 80.97% with a mean value of 79.06%. Highest milling recovery of was found in NERICA L-34 and the lowest, was observed in *Yardas* (Table 4). To obtain better quality milled rice, the knowledge of physical properties of paddy grain is necessary for modeling dynamic abrasion in a rice milling operation as well as for designing of suitable polishing systems (Mohapatra and Bal, 2004). Milling quality of rice is one of the most important criteria considered in choosing rice variety. No variety would be commercially successful unless it possessed a high turnout of whole-grain (head) rice (Webb, 1972). The milling recovery of modern rice varieties ranges from 69-73% (Biswas, *et al.*, 1992). Less than 67% milling recovery is not accepted (Dipti, *et al.*, 2003). All the tested varieties gave higher milling recovery than the reported average for modern varieties. Head rice recovery (HR) is the proportion of whole grain in milled rice. It depends on variety, grain type, cultural practice, drying temperature and of importance- effects due to moisture adsorption and desorption, as rice can be easily damaged by rapid drying or rewetting of the grains through adsorption (Webb, 1972). The broken fractions resulting from milling the paddy ranges from 13.08 (NERICA L-34) to 18.13% (*Jamila*) with a mean value of 15.45%. High broken is an undesirable trait of rice variety. Though milling qualities are appreciably high in all the tested varieties, there was a significant difference ( $p \leq 0.05$ ) among the varieties in this attributes.

Fig. 1 is the hierarchical clusters analyses groupings for the nine rice varieties studied according to their milling recovery. The varieties can be grouped into three main clusters. The first group includes two earlier released varieties (FARO 52, FARO 57), one farmer's variety (*Kwandala*) and one newly introduced variety (NERICA L-34), while the second group include one farmer's varieties (Jeep), one earlier released variety (FARO 44) and single newly released variety (NERICA L-19) and the third cluster contain two local varieties (*Jamila* and *Yardas*). The cluster analyses results indicate that both modern improved varieties and the local varieties have similarity in terms of its milling recovery.

### Conclusion and Recommendations

Physical and processing characteristics of any rice variety are important quality attribute considered when designing processing machines and also determine its marketability. Our current study has demonstrated that variation exist among the most popular rice varieties cultivated in Nigeria. Though, there was no significant variation seen in axial diameters as demonstrated by the low value of coefficient of variation. While descriptive analysis indicates standard deviation and coefficient of variation in all parameters evaluated were low indicating near uniformity and slight skewness around the mean. The cluster analyses results indicate that both modern improved varieties and the local varieties have similarity in terms of its milling recovery and record significantly high milling recovery.

Based on present data, we recommend the utilization of the data in the development of rice parboiling, milling, cleaning and grading to facilitate the mechanization of rice post production system in Nigeria to enhance the competitiveness of locally processed rice.

### Acknowledgement

Funding for this research was provided by the Canadian International Development Agency (CIDA) and the Africa Rice Center (AfricaRice) under the project '*Enhancing food security in Africa through the improvement of rice post-harvest handling, marketing and development of new rice-based product*'. The authors would also like to acknowledge National Cereals Research Institute, Badeggi and Kano State Agricultural and Rural Development Agency (KNARDA) for providing the rice varieties used in this study.

### References

- Alonge, A. F., Adigun, Y. J. (1999). Some physical and aerodynamic properties of sorghum as relates to cleaning. In proc. 21st annual Conference of the Nigerian Society of Agricultural Engineers (NSAE) at federal Polytechnic, Bauchi, Nigeria.
- Arana, A (2007). Physical and mechanical properties in rice processing. *J. Food Engg.*, 79: 137-142.
- Arora, S (1991). Physical and aerodynamic properties of rough rice (*Oryza sativa* L). *Indian J. Agric. Engg.*, 1:17-22.
- Ancheta, C.J and S.C. Aandales (1990). Total milling and head rice recovery of paddy as influenced by its physic-varietal characteristics. *Agricultural Mechanization in Asia, Africa and Latin America* 21: 50-54.
- Bergman, C.J., K.R. Bhattacharya and K. Ohtsubo (2004). Rice end use quality analysis. In: Rice: Chemistry and Technology, 3<sup>rd</sup> ed. Alaine T. Champagne (Ed.). American Association of Cereal Chemists, Inc. St. Paul,



- Minnesota, USA. Biswas, S.K., B. Banu., K.A Kabir., F. Begum and N.H Choudhury (1992). Physicochemical properties of modern and local rice varieties of Bangladesh. *Bang. Rice J.* 128-131.
- Correa, P.C., Shwanz da Silva, F., Jaren, C., Alfonso, P.C.J and I, Arana (2006). Physical and mechanical properties in rice processing. *Journal of Food Engineering* 79: 137-142.
- Danbaba, N., Anounye, J.C., Gana A.S., Abo, M.E. and Ukwungwu, M.N (2011). Grain quality characteristics of Ofada rice (*Oryza sativa* L.): Cooking and eating quality. *International Food Research Journal* 18: 619-624 (2011).
- Danbaba, N., Solomon, M.D and J.C Anounye (2012). Some Technological properties of castor seeds (*Recinus communis*) of importance in the design of its processing operations. *Academic Research International*, 2 (3): 239-245.
- Dipti, S.S., Bari, M.N and Kabir, K.A (2003). Grain quality characteristics of some Beruin rice varieties of Bangladesh. *Pakistan Journal of Biological Sciences*, 24, 1589-1594.
- Eke, C.N.O., Aseogwu, S.N and G.I. Nwandikon (2007). *Agricultural Engineering International: the CIGR E-journal of scientific research and development*. Manuscript FP 07 014 Vol. IX.
- FAOSTAT (2011). *Rice Production*. Available from <http://faostat.fao.org>.
- Ghadge, P.N., Vairagar, P.R and Prasad, K (2008). Physical properties of Chick pea split (*Cirarietinum* L.) *Agricultural Engineering International: the CIGR E-journal of scientific research and development*. Manuscript FP 07 039 Vol. X.
- Ghadge, P.N and K. Prasad (2012). Some physical properties of rice kernels: Variety PR-106. *J. Food Process. Technol.* 3(8): 175-179.
- Jain, R.K and S. Bal (1997). Properties of peal millet. *J. Agric. Engg. Res.*, 66:85-91
- Juliano, B.O (2007). *Rice Chemistry and Quality*. Philippine Rice Research Institute, Philippine.
- Kachru, R.P., Gupta R.K and A. Alam (1994). *Physico-chemical constituents and Engineering properties of food crops*. Scientific publishers, Jodhpur, India.
- Kergna, A and O. Coulibaly (2001). Etude sur la filiera de transformation du riz local au Mali: les decortiqueuses et les rizeries. Rapport Finale, Institut d'economie Rurale (IER), Ministere du Developpement Rural, Mali.
- Kunze, O.R (1985). Effect of variety and environmental factors on milling qualities of rice. In: Rice grain quality and marketing, International Rice Research Institute (IRRI), Los Bannos, Philippines.
- Luh, B.S (1980). *Rice Production and Utilization*, AVI publishing company, Inc., West Port, C.T.
- Maduako, J.N, Faborode M.O (1990). Some physical properties of cocoa pods in relation to primary processing. *Ife Journal of Technology* 2: 1-7.
- Mohsenin, N.N (1996). Physical properties of plants and animal materials. 2<sup>nd</sup> edition (revised). Gordon and Breach Scientific Publications, New york.
- Mohapatra, D and S. Bal (2004). Wear of rice in an abrasive milling operation, part 2: prediction of bulk temperature rise. *Bio-systems Engg.*, 89: 101-108.
- Morita, T and R.P. Singh (1979). Physical and thermal properties of short-grain rough rice. *Trans. ASAE*, 22: 630-636.
- NRDC (2010). National Rice Development Strategy. Federal Ministry of Agriculture and Rural Development, Abuja, Nigeria, pp 104.
- Oje, K and Ugbor, E.C (1991). Some physical properties of locus bean seed. *Journal of Agricultural Engineering Research*, 74: 213-215.
- Oje, K. 1993. Some engineering properties of Thevetia nut. *Journal of Agricultural Engineering and Technology* 1: 38-45.
- Omobuwaju, T.O., Akande, E.A and Sanni L.A (1999). Selected physical, mechanical and aerodynamic properties of African bread fruits (*Treculia Africana*) seeds. *Journal of Food Engineering*. 40: 241-246.
- Reddy B.S and A. Chakraverty (2004). Physical properties of rough and parboiled paddy. *Bio-system Engg.*, 88: 461-466.
- Saeid, M., R.A Mohammed., M.K Mohammed and T. Tavakoli (2007). Effect of de-owning and moisture content on husking characteristics of paddy in rubber-roll husker. *American-Eurasian J. Agric., and Environ. Sci.* 2 (1): 01-05.
- SAED (2000). Bilan de la collecte, la transformation et la commercialization de la production du riz a travers les rizeries et les mini-rizeries de la Vallee du Fleuve Senegal-Hivernage 1999/2000. Appui a la professionnalisation des organisations professionnelles agricoles. Direction de la Planification et du Developpement Rural, SAED/DPDR/DPE/MW.
- Steffe, J.T and R.P Singh (1980). Liquid diffusivity of rough rice components. *Trans. ASAE*, 23: 767-774
- Suthar, S.H and S.K Das (1996). Some physical properties of karingda [*Citrullus lanatus* (Thumb) Mansf] seeds. *J. Agric. Engg. Res.*, 65: 15-22.
- Simonyan, K.J., A M El-Okene and Y D Yiljep (2007). "Some Physical Properties of Samaru Sorghum 17" *Agricultural Engineering International: the CIGR Ejournal Manuscript FP 07008*. Vol. IX.
- Varnamkhasi, M.G., Hossein M., Jafari A., Rafiee S., Mohsen H and K Kamran (2007). Some Engineering Properties of Paddy (var. Sazandegi). *International Journal of Agriculture and Biology* 9 (5): 763-766.
- Webb, B.D (1972). Criteria of rice quality in the United States. In: B.O Juliano (ed.). Rice chemistry and Technology. American Association of Analytical Chemists, Minnesota, U.S.A 403-524.
- Zhout, Z., Robards K., Heliwell S and C. Blanchard (2002). Aging of stored grain: change in chemical and physical attributes. *Journal of Cereal Science* 35: 65-78.

## Tables

**Table 1:** Some physical characteristics of some local rice (paddy) varieties in Nigeria

Variety	Property	Mean	Standard deviation	Minimum value	Maximum value	Kurtosis	Skewness	CV (%)
JAMILA	Length (L)(mm)	9.47	0.58	8.60	10.5	-0.84	-0.11	6.12
	Width (W) (mm)	2.50	0.17	2.20	2.80	-0.54	0.38	6.80
	L/W Ratio	3.80	0.27	3.36	4.29	-0.80	0.15	7.11
	Thickness (T) (mm)	2.01	0.20	1.50	2.30	1.72	-0.90	9.95
	Geometric mean diameter (mm)	3.57	0.19	3.06	3.87	1.45	-0.93	5.32
	Sphericity (%)	37.77	1.44	35.17	40.57	-0.45	0.32	3.81
	Aspect ratio	26.45	1.90	23.30	29.79	-0.76	0.10	7.18
	Surface area (mm <sup>2</sup> )	40.22	4.19	29.44	46.98	1.04	-0.75	10.42
	1000-Paddy weight (g)	38.00	2.0	36.00	41.00	-1.18	0.90	5.26
	Volume (mm <sup>3</sup> )	25.43	3.85	16.27	31.77	0.47	-0.39	15.14
JEEP	Length (L)(mm)	9.60	0.44	8.70	10.20	-0.96	-0.16	4.58
	Width (W) (mm)	2.71	0.21	2.30	3.01	-0.77	-0.12	7.75
	L/W Ratio	3.56	0.31	2.89	4.25	0.73	-0.08	8.71
	Thickness (T) (mm)	2.23	0.19	1.98	2.78	1.91	1.28	8.52
	Geometric mean diameter (mm)	3.86	0.17	3.55	4.17	-1.02	0.17	4.40
	Sphericity (%)	40.30	2.34	36.46	47.93	5.09	1.69	5.81
	Aspect ratio	28.31	2.54	23.53	34.60	0.98	0.69	8.97
	Surface area (mm <sup>2</sup> )	46.96	4.20	39.68	54.62	-1.03	0.23	8.94
	1000-Paddy weight (g)	29.50	0.84	29.00	31.00	1.43	1.54	2.85
	Volume (mm <sup>3</sup> )	30.84	4.13	23.93	28.21	-1.08	0.23	13.39
YARDAS	Length (L)(mm)	9.13	0.60	8.30	10.40	-0.36	0.76	6.57
	Width (W) (mm)	2.65	0.12	2.50	3.00	2.05	0.29	4.53
	L/W Ratio	3.45	0.29	2.93	4.08	0.10	0.51	8.41
	Thickness (T) (mm)	2.36	0.20	1.91	2.67	0.00	-0.45	8.47
	Geometric mean diameter (mm)	3.84	0.16	3.59	4.26	0.88	0.64	4.17
	Sphericity (%)	42.22	2.56	35.76	46.87	1.04	-0.43	6.06
	Aspect ratio	29.20	2.43	24.51	34.09	-0.10	-0.08	8.32
	Surface area (mm <sup>2</sup> )	46.49	4.02	40.55	57.01	1.18	0.78	8.65
	1000-Paddy weight (g)	27.22	0.44	27.00	28.00	4.80	2.18	1.62
	Volume (mm <sup>3</sup> )	30.15	3.88	24.54	40.75	1.71	0.78	12.87
KWANDAL A	Length (L)(mm)	9.28	0.54	8.20	10.20	0.26	-0.07	5.82
	Width (W) (mm)	2.66	0.17	2.40	3.00	-0.29	0.43	6.39
	L/W Ratio	3.50	0.29	3.30	4.17	0.25	0.32	8.29
	Thickness (T) (mm)	2.38	0.15	2.09	2.67	0.60	-0.03	6.30
	Geometric mean diameter (mm)	3.88	0.18	3.51	4.15	-0.63	0.20	4.64
	Sphericity (%)	41.36	2.17	37.41	45.66	-0.18	0.40	5.25
	Aspect ratio	28.74	2.37	24.00	32.97	-0.14	0.13	8.25
	Surface area (mm <sup>2</sup> )	46.23	4.31	38.79	54.16	-0.62	0.29	9.32
	1000-Paddy weight (g)	30.20	0.84	29.00	31.00	-0.61	-0.51	2.78
	Volume (mm <sup>3</sup> )	31.00	4.40	23.76	39.31	-0.61	0.39	14.19

**Table 2:** Some physical characteristics of some popular improved rice (paddy) varieties in Nigeria

Variety	Property	Mean	Standard deviation	Minimum value	Maximum value	Kurtosis	Skewness	CV (%)
FARO 44	Length (L)(mm)	9.41	0.40	8.80	10.00	-1.49	0.15	4.25
	Width (W) (mm)	2.47	0.19	2.10	2.80	-0.76	-0.21	7.69
	L/W Ratio	3.83	0.27	3.41	4.32	-0.77	0.22	7.05
	Thickness (T) (mm)	2.20	0.18	2.60	2.61	-0.38	0.57	8.18
	Geometric mean diameter (mm)	3.70	0.21	3.30	4.08	-1.11	0.14	5.68
	Sphericity (%)	39.34	1.91	36.33	43.29	-0.23	0.30	4.86
	Aspect ratio	20.23	1.84	23.16	29.35	-0.80	0.02	9.10
	Surface area (mm <sup>2</sup> )	43.15	4.90	35.04	52.22	-0.08	0.21	11.36
	1000-Paddy weight (g)	26.60	0.55	26.00	27.00	-3.33	-0.61	2.07
	Volume (mm <sup>3</sup> )	26.99	4.59	19.59	35.83	-1.01	0.27	17.01
FARO 52	Length (L)(mm)	10.41	0.35	9.60	11.00	0.88	-0.68	3.36
	Width (W) (mm)	2.53	0.15	2.30	2.90	0.33	0.73	5.93
	L/W Ratio	4.12	0.27	3.52	4.42	-0.48	-0.75	6.55
	Thickness (T) (mm)	2.25	0.15	2.00	2.50	-1.15	0.14	6.67
	Geometric mean diameter (mm)	3.89	0.15	3.59	4.18	-0.14	0.20	3.86
	Sphericity (%)	37.43	1.73	35.13	40.82	-0.82	0.49	4.62
	Aspect ratio	24.35	1.67	22.64	28.43	0.07	0.94	6.86
	Surface area (mm <sup>2</sup> )	47.72	3.70	40.58	55.01	-0.14	0.29	7.75
	1000-Paddy weight (g)	27.40	0.55	27.00	28.00	-3.33	0.61	2.01
	Volume (mm <sup>3</sup> )	31.33	3.65	24.46	38.61	-0.09	0.39	11.65
FARO 57	Length (L)(mm)	9.67	0.71	8.80	10.90	-1.08	0.52	7.34
	Width (W) (mm)	2.40	0.16	2.10	2.80	1.15	0.73	6.67
	L/W Ratio	4.04	0.37	3.26	4.61	-0.46	-0.37	9.16
	Thickness (T) (mm)	2.17	0.18	1.89	2.56	-0.89	-0.37	8.29
	Geometric mean diameter (mm)	3.69	0.20	3.34	4.12	-0.57	0.14	5.42
	Sphericity (%)	38.23	2.46	33.62	43.13	-0.26	0.32	6.43
	Aspect ratio	24.98	2.38	21.70	30.68	0.14	0.75	9.53
	Surface area (mm <sup>2</sup> )	42.82	4.71	35.04	53.34	-0.42	0.25	11.00
	1000-Paddy weight (g)	37.80	1.30	36.00	39.00	-1.49	-0.54	7.34
	Volume (mm <sup>3</sup> )	26.67	4.38	19.60	36.88	-0.19	0.35	16.42

FARO = Federal Agricultural Research Oryza

**Table 3:** Some physical characteristics of some newly released improved rice (paddy) varieties in Nigeria

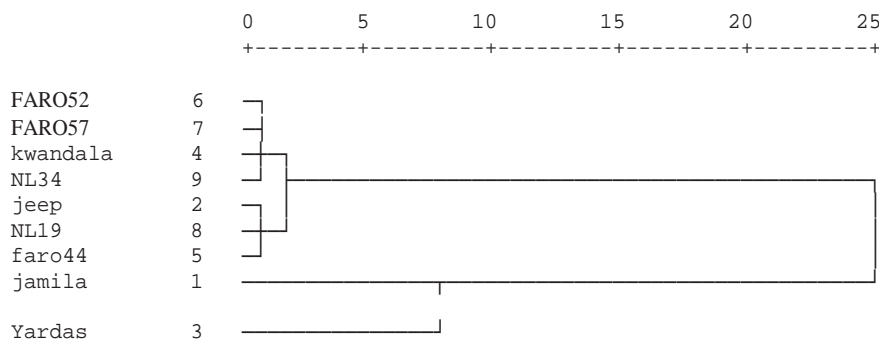
Variety	Property	Mean	Standard deviation	Minimum value	Maximum value	Kurtosis	Skewness	CV (%)
NERICA L-19	Length (L)(mm)	8.52	0.40	7.70	9.00	-0.78	-0.50	4.69
	Width (W) (mm)	2.66	0.14	2.40	2.90	0.02	-0.11	5.26
	L/W Ratio	3.21	0.23	2.89	3.70	-0.73	0.54	7.17
	Thickness (T) (mm)	2.43	0.26	2.00	3.40	9.50	2.49	10.70
	Geometric mean diameter (mm)	3.79	0.19	3.43	4.38	4.49	1.25	5.01
	Sphericity (%)	44.60	2.46	40.49	49.28	-0.85	-0.15	5.52
	Aspect ratio	31.29	2.20	26.97	34.57	-0.01	-0.34	7.03
	Surface area (mm <sup>2</sup> )	45.34	4.61	36.85	60.43	5.36	1.54	10.17
	1000-Paddy weight (g)	32.61	1.67	30.00	34.00	0.54	-1.09	5.12
Volume (mm <sup>3</sup> )	29.05	4.61	21.30	44.82	6.54	1.89	15.87	
NERICA L-34	Length (L)(mm)	9.64	0.59	8.40	10.50	-0.27	-0.41	6.12
	Width (W) (mm)	2.54	0.12	2.30	2.70	-0.88	-0.19	4.72
	L/W Ratio	3.80	0.30	3.19	4.20	-0.80	-0.55	7.89
	Thickness (T) (mm)	2.21	0.14	1.98	2.40	-1.15	-0.15	6.33
	Geometric mean diameter (mm)	3.77	0.15	3.48	4.01	-0.84	-0.16	3.98
	Sphericity (%)	39.24	2.21	26.06	43.92	0.03	0.83	5.63
	Aspect ratio	26.47	2.19	23.76	31.39	-0.41	0.77	8.27
	Surface area (mm <sup>2</sup> )	44.83	3.51	38.06	50.55	-0.88	-0.09	7.83
	1000-Paddy weight (g)	36.60	0.89	36.00	38.00	0.31	1.26	2.43
Volume (mm <sup>3</sup> )	28.57	3.30	22.28	34.07	-0.91	-0.03	11.55	

NERICA = New Rice for Africa

**Table 4:** Milling characteristics of some popular rice varieties in Nigeria

Variety	Milling properties				
	Hull (%)	Head rice recovery (%)	Broken (%)	Milling degree (%)	Milling recovery (%)
<i>Jamila</i>	18.26 <sup>e</sup>	67.99 <sup>d</sup>	18.13 <sup>b</sup>	3.13 <sup>ab</sup>	76.84 <sup>d</sup>
<i>Jeep</i>	19.63 <sup>c</sup>	70.16 <sup>a</sup>	16.16 <sup>c</sup>	2.88 <sup>b</sup>	79.59 <sup>bc</sup>
<i>Yardas</i>	20.12 <sup>b</sup>	66.42 <sup>e</sup>	15.13 <sup>e</sup>	3.18 <sup>a</sup>	74.27 <sup>e</sup>
<i>Kwandala</i>	18.14 <sup>e</sup>	69.15 <sup>bc</sup>	10.67 <sup>h</sup>	2.19 <sup>cd</sup>	80.26 <sup>ab</sup>
FARO 44	19.74 <sup>c</sup>	70.05 <sup>a</sup>	19.27 <sup>a</sup>	2.34 <sup>c</sup>	79.27 <sup>c</sup>
FARO 52	19.41 <sup>d</sup>	68.37 <sup>cd</sup>	15.72 <sup>d</sup>	2.13 <sup>cd</sup>	80.41 <sup>a</sup>
FARO 57	18.11 <sup>e</sup>	69.16 <sup>bc</sup>	14.46 <sup>f</sup>	2.04 <sup>d</sup>	80.45 <sup>a</sup>
NERICA L-19	21.01 <sup>a</sup>	70.07 <sup>a</sup>	14.60 <sup>f</sup>	2.90 <sup>a</sup>	79.51 <sup>bc</sup>
NERICA L-34	18.17 <sup>e</sup>	69.37 <sup>ab</sup>	13.08 <sup>g</sup>	3.03 <sup>ab</sup>	80.97 <sup>a</sup>
Mean	19.18	69.97	15.45	2.65	79.06
CV (%)	0.60	0.60	0.80	5.10	0.50
LSD (5%)	0.209	0.759	0.214	0.235	0.722

Data are mean value of three readings. Mean in the same column having different letter are significantly different at 5% level of probability.



**Figure 1:** Hierarchical cluster showing vertical groupings of the rice varieties according to percentage milling recovery

## QUALITY ASSESSMENT OF BIODIESEL PRODUCED FROM AFTER FRY WASTE PALM KERNEL OIL (PKO)

**Eze, S.O.**, Abia State University, Nigeria  
E-mail: sundayoeze@yahoo.com

**Ngadi, M. O.**, McGill University, Quebec, Canada

**Alakali, J.S.**, Federal University of Agriculture, Markurdi, Benue State, Nigeria

**Odinaka, C.J.**, Abia State University, Nigeria

**Cite this Paper:** Eze, S. O., Ngadi, M. O., Alakali, J. S. & Odinaka, C. J. (2013). Quality Assessment of Biodiesel Produced From After Fry Waste Palm Kernel Oil (PKO). *European Journal of Natural and Applied Sciences*, 1(1), 38-46.

### ABSTRACT

*Biodiesel, produced from waste palm kernel oil (PKO) collected after 3, 6, and 9 hours of frying were analysed for biodiesel quality characteristics in the continued effort to find suitable sources of biodiesel and as an environmental cleaning process. The biodiesel yield of the fresh palm kernel oil was 90%, while the after fry waste oils used for 3, 6 and 9 hours yielded 87.56%, 88.75% and 87.50% respectively. The kinematic viscosity at 40°C and 100°C were between 4.70-7.54 and 1.64- 2.49 cSt respectively. Other results included specific gravity (0.881-0.894g/cm<sup>3</sup>), pour point (3.0-11.0°C, flash point (76-124°C), water content (0.67-2.5 % ), API gravity (102.58-124.8), Cetane index(60.6-67.4), aniline point (10.0°C-15.0°C), and average boiling point (90°C-110°C. The colour of the samples were light yellow, light brown, brown and dark brown for the biodiesel produced from the fresh palm kernel oil and the waste oils recovered after 3,6 and 9 hours of use respectively. The acid values were 2.24-2.58, iodine value was 3.0-8.9, peroxide value was 2.0 -8.0 and saponification value was 268-335. The results compared fairly with reference quality specifications for diesel fuel standards showing that the waste oil is useful for biodiesel production.*

**Keywords:** Biodiesel, waste, palm kernel oil, quality.

### INTRODUCTION

Biofuels, which are fuels obtained from plants and animal products have continued to attract a lot of attention. This is because of the continued need for alternative, sustainable, cleaner and renewable energy which fossil fuels fall short of. Fossil fuels which are hydrocarbons primarily- coal, petroleum and natural gas have been the world's major source of energy for a long time. However, they are fraught with some problems such as their polluting effects as they contribute significantly to the greenhouse effect through the emission of carbon (IV) oxide, CO<sub>2</sub>, and other toxic gases including sulphur oxides and nitrogen oxides which result to acid precipitation, as well as particulate matter in the form of unburnt carbon. The high cost of fossil fuels and the possibility of its depletion is also another factor that favour biofuels. Biofuels have the advantage of being biodegradable, less polluting and renewable (Falcon and Siekmam, 1997).



Biodiesel is a fuel obtained from a process of trans-esterification in which an ester is converted to another ester by reacting it with an alcohol, a carboxylic acid or another ester (Fukuda *et al*, 2001). It is a mixture of monoalkyl esters of long chain fatty acids derived from biomass-especially vegetable oils or animal fat which has been a domestic renewable fuel for diesel engines and meets the specifications of ASTM .D6751(Guo and Leung,2003). The use of biodiesel is preferred because it is less viscous, and forms less particulate matter on combustion unlike the direct use of the straight vegetable oils.

The use of the straight vegetable oils, although energetically favourable is problematic because of high viscosity (about 11-12 times higher than diesel fuel). It also does not burn well and completely and forms deposits in fuel injector engines as well as the formation of acrolein through the thermal decomposition of glycerol (Freedman and Bagby, 1990). The problem of high viscosity of straight run vegetable oil is checked by transesterification, pyrolysis (catalytic or thermal cracking), and dilution with conventional petroleum diesel as blends and microemulsification. It is true that the processes of microemulsification , and thermal cracking has been proven ineffective and blending does not decrease the use of fossil fuels. (Meka,Tripath and Singh,2007). Transesterification is therefore the current favoured process (Clark *et al*, 1984).

Biodiesel is produced from a wide variety of vegetable oils including virgin oil feedstocks such as palm, soybean, rapeseed, mustard, peanut, coconut, olive, Jatropha, flax, sunflower, and cotton seed oils. Others are rubber seed, grape seed, hazelnut, linseed safflower, sesame, and corn and rice bran oils (Paynich, (2007). The use of these vegetable oils is considered a cheap source especially when they have been used for frying. The use of fry vegetable oils is also useful environmentally as a means of disposing off restaurant greases or domestic fry oil wastes and reducing the cost of waste treatment. (De Poola *et al*, 2009). Biodiesel is also produced from animal fats (Hilbert, et al. 2005) such as tallow, lard, yellow grease and chicken fat as well as the by-products of production of omega-3-fatty acids from fish oil.

The production of biodiesel from algae has received a lot of publicity (Chrsti, 2007, Christi, 2008, Greenwell *et al*, 2010, Gordon, and Polle, 2007). The transesterification process usually involves the use of short chain alcohols such as methyl or ethyl alcohol and a catalyst is normally used to speed up the reaction. The catalyst can be a base or acid or enzymes such as lipase and many others (Shimada *et al*, 2002).

Acid catalysis is viewed as being too slow kinetically. The use of bases as catalysts has the problem of substrate loss through soap formation, high viscosity, formation of gels and the difficulty in separation of glycerol which is formed as a byproduct. The alternative enzymatic catalysis has the advantage of the possibility of carrying out the reactions at low temperatures ( Paynich, 2007). Also the enzymes such as lipase can catalyse the transesterification of free fatty acid (Kumari *et al*,2007; Kaieda *et al*, 2001) .The other advantage is higher yields of esters comparatively, and the recovery and reuse of enzymes. However, the cost of the enzymatic process is still too high (Hernández-Martín and Otero, 2008).

In this paper we have used waste palm kernel fry oil and the base catalysis process which still seem to be the best practical option. We have examined the diesel quality characteristics of the biodiesel produced from this waste oil at different stages of use in frying relative to the biodiesel produced from the fresh palm kernel oil.

## MATERIALS AND METHODS

### Sample Collection and Preparation

The fresh palm kernel oils were bought from Aniuzo International Limited, Palm Kernel Oil Mills Division, Emene Industrial Layout, Emene, near Enugu, Enugu State, Nigeria. The chemicals- Methanol, BDH, Chloroform, CVL, sodium hydroxide (BDH), Acetic acid, (Chawell Heath Essex, England), sodium thiosulphate (Hopkins & William Ltd), Wijis Solution (Sigma-Aldrich) and Potassium iodide (M&B) were of analytical grade and used without further purification. The fresh oil sample was used in frying potatoes and the used oils collected after 3, 6, and 9 hours of frying and stored in clean bottles for the biodiesel production.

### Experimental Methods

The fresh samples as well as the fry oils were used for biodiesel production. The transesterification reaction was run in a batch reactor submerged in a water bath set at 65° C and containing 1 litre of water and allowed to equilibrate at this temperature. Four hundred milliliters (400ml) of oil sample was put in the reactor and allowed to steam for 30 minutes. Then 3.6g sodium hydroxide was weighed and put in a beaker and 76.78ml of methanol added to the beaker containing 1% NaOH and the mixture stirred vigorously to dissolve the NaOH pellets. The solution was poured into the reactor containing the oil sample. The reaction was allowed to run for 40 minutes with continuous stirring. The contents were poured into a separator funnel for the separation of glycerol from the biodiesel. The same procedure was applied to the fresh sample and the fry oils used for 3, 6 and 9 hours.

### Sample Analysis

The biodiesel samples produced were analysed for diesel quality characteristics using standard methods of analysis. These include kinematic viscosity and Flash point, (ASTM, 1999), pour point (ASTM, 2006), as well as saponification value, iodine value, and peroxide value, (Boerlage and Broeze, 1990), acid value (ASTM, 2004) carbon residue, (ASTM 01149, 2009), moisture content, (Wick, 1985), aniline point, (ASTM 1994), boiling point, (Schrock, 1984), smoke point (Schrock, 1984), colour, (ASTM 1990), and cetane number and cetane index (Pischinger and Falcon, 1993, ASTM, 1990, ASTM, 1997). The ASTM D613 test for determining cetane index is time consuming and expensive. The calculated cetane index ASTM D967 or D4737 is often substituted for the cetane number. The calculated cetane index is decided from the fuel's boiling range and density (Hobson, 1984, Boerlage and Broeze, 1994) It is calculated as Cetane Index (CI) = API gravity x Aniline Point / 100

Where API gravity = Mid boiling point / Specific gravity.

## RESULTS AND DISCUSSION

### Percentage Yield

The result of the percentage yield of biodiesel samples produced is shown in table 1 below

**Table 1: Percentage Yields of Biodiesel Samples Produced**

Sample	Catalyst	Reaction Temperature (° C)	Reaction Time (mins)	% Weight of Catalyst	Alcohol-Oil Molar ratio	Biodiesel Yield (%)
A (Fresh PKO)	NaOH	65	40	1	6:1	90.00

<b>B (Waste fry PKO used for 3 hours)</b>	NaOH	65	40	1	6:1	87.56
<b>C (Waste fry PKO used for 6 hours)</b>	NaOH	65	40	1	6:1	88.75
<b>D (Waste fry PKO used for 9 hours)</b>	NaOH	65	40	1	6:1	87.5

From table 1 above, the yield (87.5-88.75%) of biodiesel from the waste oils between quite encouraging although it was lower than the yield from the fresh oil sample. This is higher than the results of Laita and Sukuny, 2004 who got 72% with crude palm kernel oil but lower than 96% obtained by Chitra *et al*, 2005 from alkali-catalysed transesterification of *Jatropha curcus* oil.

### Biodiesel Quality Characteristics

The results of the diesel quality characteristic tests for the fresh and waste fry oils are shown in table 2 below.

**Table 2: Biodiesel Quality characteristics of the Biodiesel from the Waste Fry Oils and from the Fresh oil**

Sample	A	B	C	D	ASTM Diesel Standards
<b>Kinematic viscosity at 40 °C (cSt)</b>	4.70	4.36	4.64	7.54	1.60-6.50
<b>Kinematic viscosity at 100 °C (cSt)</b>	1.64	1.81	1.78	2.49	1.60-6.50
<b>Specific gravity (g/cm<sup>3</sup>)</b>	0.897	0.875	0.881	0.894	0.820-0.870
<b>Pour point(°C)</b>	3.00	3.00	4.00	11.00	
<b>Flash point(°C)</b>	124.00	112.00	76.00	78.00	66 .00
<b>Water content (%)</b>	0.83	0.67	1.40	2.40	0.05 max
<b>API gravity</b>	117.00	102.58	124.8	114..00	150.00
<b>Cetane Index</b>	65.50	60.60	67.40	57.00	47.00

### Viscosity

The flow properties of a fuel are important for its proper functioning at different temperatures. The kinematic viscosity of the biodiesels was within acceptable standard of ASTM and Engine Manufacturers Association (EMA).

### Specific Gravity and API Gravity

The specific gravities of the biodiesels produced from the waste fry oils which were between 0.875 and 0.897 show that they can be satisfactorily used to operate diesel engines as they are not too light or too heavy. They also compare favourably to the recommended 0.76-0.88 for diesel (Weiksner *et al*, 2001) for diesel fuel as well as the guideline for pure biodiesel (B100) according to ONORM C1190 used in Austria. The API gravity is usually calculated from the density and the values of 102.58-124.0.

### Pour Point

The pour point is usually taken as the lowest temperature at which the fuel has the ability to flow or the lowest temperature at which a fuel can be pumped from the fuel tank to the combustion chamber. Its importance is significant in low temperature regions. The pour points of the biodiesel sample lie between 3 and 11°C which are acceptable. Pour point standards seem not to be indicated by many standard bodies. However, Official Journal of France gave the value of <10 °C in summer for vegetable oil methyl esters (VOME) (Lele, 2010).

### Flash point

The flash point is the lowest temperature at which a fuel ignites. Sample A has the highest flash point followed by B while samples C and D values are lower. The values obtained for the biodiesel samples (for the fresh and used fry oils fall between 76 and 124 and are within the ASTM standard of > 130°C This lowest temperature at which a fuel ignites for the biodiesel samples obtained from the fresh and used fry oils between 76-124 was within the ASTM standard of >130°C (min) and 150°C (max).

### Cetane Index/Cetane Number

The Cetane Number is a measure of the ignition quality of diesel fuel. The higher the cetane number, the easier it is to start a standard (direct injection) diesel engine. It is a measure of the % by volume of cetane (hexadecane) in a combustible mixture (containing cetane and 1-methylnaphthalene) whose ignition characteristics match those of diesel fuel being tested. The calculated cetane indices of the biodiesel samples which were between 47.00 and 67.00 are not far from the standard of 47 minimum by ASTM or 51 minimum by IS 15607:2005. Chemical characteristics of the biodiesel sample are shown in table 3 below.

**Table 3: Chemical Characteristics of the Biodiesels**

Sample	A	B	C	D	ASTM/EN Diesel standards
Acid Value Mg KOH/g	2.24	2.25	2.24	2.58	0.80/0.50 (max)

<b>Iodine Value</b>	3.00	3.6	8.90	8.40	<120
<b>Peroxide Value (ppm)</b>	2.00	4.00	4.00	8.00	n.s
<b>Saponification Value</b>	319.00	335.00	273.00	268.00	n.s
<b>Aniline point(°C)</b>	13.30	15.00	12.20	10.00	
<b>Boiling point (°C)</b>	105.00	90.00	110.00	102.00	
<b>Colour</b>	Light yellow	Light brown	Brown	Dark brown	

### Acid Value

The acid value of the diesel samples (2.24-2.58) is higher than the ASTM standard of 0.80mgKOH/g as well as the EN standard of 0.50mgKOH/g. This may be attributed to the increase in free fatty acid through chain pyrolysis during frying and unconverted acid during transesterification. This could also explain the lower ester yields (87.5-90.0%) instead of the expected standard yield of 96.5% for biodiesels by Indian Standards(IS). However, higher acid values (10.568mgKOH/g) have been reported for biodiesel from *Jatropha* using  $\text{KNO}_3/\text{Al}_2\text{O}_3$  catalyst ( Vyas, *et al*,2009 ).

### Iodine Value

The iodine values are significantly low (3.0-8.4) so far below the specification of less than 120. The reason could be explained by nature of the raw material- used (palm kernel fry oils) and the type of fatty acids it contains. The typical composition of palm kernel oil is shown in table 4.

**Table4: Composition of Palm Kernel Oil**

<b>Fatty Acid</b>		<b>Number of carbon atoms</b>	<b>Type</b>	<b>Percentage composition</b>
<b>Trivial Name</b>	<b>Systematic Name</b>			
Lauric acid	Dodecanoic acid	C <sub>12</sub>	Saturated	48.2
Myristic acid	Tetradecanoic acid	C <sub>14</sub>	Saturated	16.2
Palmitic acid	Hexadecanoic acid	C <sub>16</sub>	Saturated	8.4
Capric acid	Decanoic acid	C <sub>10</sub>	Saturated	3.4
Caprylic	Octanoic acid	C <sub>8</sub>	Saturated	3.3
Stearic acid	Octadecanoic acid	C <sub>18</sub>	Saturated	15.3
Oleic acid	Cis-9-octadecanoic acid	C <sub>18</sub>	Monounsaturated	2.3
Others/unknown				0.40



The saturated/unsaturated fatty acid in palm kernel oil is 94.8:2.3 from the composition shown in the table. This accounts for the very low iodine value since iodine value is a measure of degree of unsaturation in an organic compound. This also indicates that palm kernel oil would not be very good cooking oil. Its application for biodiesel production is better as well as other industrial applications such as in soap making.

### **Peroxide Value**

The peroxide value is not usually considered as very suitable in evaluating oxidation and therefore the stability of biodiesel. It is also not specified in fuel standards. However, they can be of concern as they are formed as a result of biofuel oxidation and ultimately decompose to aldehydes, acids and polymers. These products can affect the plastics and elastomers in the fuel system. The peroxide value of the biodiesel from PKO and its waste fry oils ranged from 2-8 meqO<sub>2</sub>. Higher values from 10- over 100 ppm have been reported (MoCormick, *et al*)

### **Saponification Value**

The saponification values of 268-335 for the biodiesel samples are very high as would be expected because of their production from highly saturated oil as PKO. This is so far higher than (52-62mgKOH/g) reported by Polak, (2004) for biodiesel produced from a more unsaturated oil such as rape seed oil.

### **Aniline Point**

The aniline point of the biodiesel samples ranged from 10-15°C. The aniline point temperature is important because of its relationship to the cetane index which is important in assessing diesel quality as indicated in the equation:

$$\text{Cetane Index} = \text{Aniline Point (}^{\circ}\text{C)} \times \text{API Gravity}/100$$

The aniline point is the lowest temperature at which equal volume of aniline and the oil under test form a single phase. The aniline point (AP) is related to the amount and type of aromatic hydrocarbons in oil sample. If the AP is low, it means there is high aromatic content in the oil while a high AP is indicative of low aromatic content. High AP will also result in high cetane index and vice versa.

### **Boiling Point**

The boiling point is not usually as useful as the distillation profile used in calculating the cetane index. The boiling point of the samples ranged from 90°C to 110°C. It seems from the result there may be chemical transformations during the frying process leading to the variations in the boiling points of the samples.

### **Colour**

The colour of the biodiesel samples produced from the used oil deepen with the number of hours of usage from light yellow for the unused oil to light brown, brown and dark brown for the samples produced from oils used for 3, 6 and 9 hours respectively. The colour can be

improved by bleaching the used oil before biodiesel production. This is likely to increase the cost of production.

## CONCLUSION

The use of waste fry palm kernel oil for the production of biodiesel is promising. The biodiesel properties of the samples did not deviate significantly from expected values. So fresh palm kernel oil as well as its waste fry oil can be used advantageously as a raw material for biodiesel production.

## ACKNOWLEDGEMENT

The authors are grateful to the Director, Product Development Authority (PRODA), Enugu, Nigeria and the Head, R& D, for their assistance in the use of some of their equipment.

## REFERENCES

- Annual Book of ASTM Standard Test Methods (1994) Aniline Value and smoke point, *Proceedings of Annual book of American Society of for Testing and materials*, 15, 259
- Annual Book of ASTM Standard Test Methods (1997) Specification for Diesel fuel Properties, *Proceedings of Annual book of American Society of for Testing and materials*, 6, 110-113
- Annual Book of ASTM Standard Test Methods (1999) Aniline Value and smoke point, *Proceedings of Annual book of American Society of for Testing and materials*, 2, 70.
- Annual Book of ASTM Standard Test Methods (1999) Determination of Flash Point, *Proceedings of Annual book of American Society of for Testing and materials*, 5, 40-41.
- Annual Book of ASTM Standard Test Methods (2006) Determination of Pour Point, *Proceedings of Annual book of American Society of for Testing and materials*, 2, 32-34
- Annual Book of ASTM Standard Test Methods (2004) Acid Value, *Proceedings of Annual book of American Society of for Testing and materials*, 2, 70.
- Annual Book of ASTM Standard Test Methods (2004) Acid Value, *Proceedings of Annual book of American Society of for Testing and materials*,
- Boerlage, G. D. and Broeze, J.J. (1990) Determination of saponification , Iodine and Peroxide Values, *Progress report of Volunteer Group for Fuel Research Society of Automotive Engineers*, 21, 289-305.
- Boerlage, G. D. and Broeze, J.J. (1994) Determination of Cetane number, *Progress report of Volunteer Group for Fuel Research Society of Automotive Engineers*, 5, 231-232.
- Clarck, S. J., Wangner, L., Shrock, M.D. and Piennarr, P.G. (1984) Methyl and ethyl soybean esters as renewable fuels for diesel engines. *J. Am. Oil Chem. Soc.* 61, no 10 :1632-1638.
- Chitra, P., Venkatachalam, P. and Sampathrajan, A. (2005) *Optimization of Experimental Conditions for Biodiesel Production from Alkali-catalysed Transesterification of Jatropha*.
- Curcas Oil, Department of Bio-Energy, ACE&RI, Tamil Nadu Agricultural University, Coimbatore-641 003, Tamil Nadu, India, *Energy for Sustainable Development*, 2005, IX, No. 3, September.
- Christi, Y (2007) Biodiesel production from microalgae, *Biotechnol. Adv.*, 25, 294-306.
- Christi Y. (2008) Response to Rejoinder: Do biofuels from microalgae beat biofuels from

- terrestrial plants, *Trends Biotechnol.* 26, 351-352.
- De Poola, M.G., Ricca, E., Calabro, V., Curcio, S., Iorio, G., (2009) Factor analysis of transesterification reaction of waste oil for biodiesel production. *Bioresource Technology*, 100: 5126-5133.
- Falcon, A.M. and Siekmam, M.N. (1997) Benefits of Biodiesel, *JAACS*, 20: 21.
- Freedman, B. and Bagby, M.O. (1990) Problems of straight vegetable oils, *Journal of American oil Chemists Society*, 12, 48.
- Fukuda H., Kondo A. and Noda H. (2001) Biodiesel fuel production by transesterification of oils. *J. Biosci. Bioeng.* 92, 405-416
- Guo, V.L. and Leung, Y. C. (2003) Biodiesel as a Domestic Renewable Fuel. *Proceedings of Macro Review Engineering Society*, 40, 421-422.
- Hernández-Martín E., Otero E.C. (2008) Different enzyme requirements for the synthesis of biodiesel: Novozym 435 and Lipozyme TL IM, *Bioresource Technol.* 99, 277-286
- Hilbert, T., Mittelbach, M., Schmidt, E. (2005) Animal fats perform well in Biodiesel, *5th International Colloquium on FUELS in Germany held January 12-13*.
- Hobson, G. D. (1984) *Standard Test Methods for Cetane number of diesel fuel oil*. Modern Petroleum Technology, John Wiley and sons.
- Kaieda M., Samukawa T., Kondo A. and Fukuda, H. (2001) Effect of methanol and water contents on production of biodiesel fuel from plant oil catalyzed by various lipases in a solvent-free system. *J. Biosci. Bioeng.* 91, 12-15
- Kumari V., Shah S. and Gupta M.N. (2007) Preparation of Biodiesel by Lipase-Catalyzed Transesterification of High Free Fatty Acid Containing Oil from *Madhuca Indica*. *Energy Fuels* 21, 368-372
- Lele, S. (2010) Biodiesel in India. *Indian Green Energy Awareness Centre Report*, 1, 1.
- Gordon J. M., Polle J. E. (2007) Ultrahigh bioproductivity from algae. *Appl. Microbiol. Biotechnol.* 76, 969-97.
- Greenwell, H. C., Laurens, L.M. L., Shields, R.S., Lovitt, R. W. and Flynn, K.J. (2010) Placing microalgae on the biofuel priority list: A review of the technological challenge. *J. R. Soc. Interface* 46, 703-726.
- Meka, P.K., Tripath, V. and Singh, R.P. (2007) Synthesis of Biodiesel from Safflower oil using various reaction parameters. *J. Oleo. Sci.* 56, 9-12.
- Paynich, M. (2007) Transesterification of vegetable oils to produce Biodiesel fuels. *MMG 445, Basic Biotechnology eJournal*, 3:57-61.
- Pischinger, G. H. and Falcon, A.M. (1993) *Cetane Number*. Cetane Number Engine Testing Comparison to calculated Cetane Index Final Report to the National Biodiesel Board 20, 755-756.
- Shimada, Y. Watanabe, A. Sugihara and Y. Tominaga (2002) Enzymatic alcoholysis for biodiesel fuel production and application of the reaction to oil processing. *J Mol Catal. B: Enzym* 17, pp. 133-142
- Shrock, S. I. (1984) *Determination of Smoke Point. Biodiesel Fuel Properties and process Technology*. Applied Science Publishers, London, p. 801.
- Vyas, A.P., Subrahmanyam, N., Patel, P.A. (2009) Production of Biodiesel through transesterification of *Jatropha* oil using  $\text{KNO}_3/\text{Al}_2\text{O}_3$  solid catalyst. *Fuel*, 88 (4):625-628.
- Weiksner, J. M., Crump, S. L. and White, T. L. (2003) *Understanding Biodiesel fuel quality and Performance*. Department of Energy, US Department of Commerce, National Technical Information Service, Energy Citation Data Base, Springfield.
- Wick, W. E. (1985) Water Content Determination. *Proceedings of Macro review Engineering Society*, 10, 225





Contents lists available at SciVerse ScienceDirect

Food Control

journal homepage: [www.elsevier.com/locate/foodcont](http://www.elsevier.com/locate/foodcont)

## Prediction of pork marbling scores using pattern analysis techniques

H. Huang<sup>a</sup>, L. Liu<sup>a</sup>, M.O. Ngadi<sup>a,\*</sup>, C. Gariépy<sup>b</sup>

<sup>a</sup> Department of Bioresource Engineering, McGill University, Macdonald Campus, 21, 111 Lakeshore Road, Ste-Anne-de-Bellevue, Quebec, Canada H9X 3V9

<sup>b</sup> Agriculture and Agri-food Canada, 3600 Cassavant West, Saint-Hyacinthe, Quebec, Canada J2S 8E3

### ARTICLE INFO

#### Article history:

Received 17 April 2012

Received in revised form

17 September 2012

Accepted 25 September 2012

#### Keywords:

Pork

Marbling

Wide line detector

Multiple linear regression

Grey-level co-occurrence matrix

### ABSTRACT

Marbling is an important technical quality attribute of pork. Its assessment usually corresponds to a subjective score being given by trained panelists based on the marbling standards charts. The purpose of this study was to investigate objective determination of pork marbling using pattern analysis techniques. A line pattern recognition technique called the wide line detector (WLD) and a texture extraction technique based on an improved grey-level co-occurrence matrix (GLCM) were employed and compared. Fifty three fresh pork loin chops from the longissimus dorsi (LD) muscle were collected and their marbling scores were assessed in a plant. Red-Green-Blue (RGB) images of these chops were acquired using a digital camera. The loin eye area was selected as the region of interest (ROI) of the pork images. Marbling was extracted from the ROI by either GLCM or WLD. Proportion of marbling (**PM**) obtained from WLD or image texture measurement from GLCM (**GI**) was formulated as indices of the marbling score. Linear regressions based on the **PM** and **GI** were carried out at the red, green, and blue channels as well as the combined RGB channels. The results of WLD and GLCM based models showed the effectiveness of pattern analysis techniques for pork marbling assessment. The comparison indicated that the WLD based models had stronger predictive ability for pork marbling score determination than GLCM. The green channel was demonstrated to have the best explanatory for pork marbling assessment no matter which pattern analysis technique used. High correlation coefficients of calibration and validation ( $R_c = 0.94$ ,  $R_v = 0.94$ ) of the WLD based linear model at the green channel strongly indicated the great potential of pattern analysis techniques especially the line pattern recognition methods for the accurate and real-time evaluation of pork marbling.

© 2012 Elsevier Ltd. All rights reserved.

### 1. Introduction

Marbling is a visual attribute of meat that affects the acceptability and palatability of pork in market and is defined as the amount and spatial distribution of the visible fat within the longissimus dorsi (LD) muscle. In the pork industry, the evaluation of marbling mainly relies on subjective comparison with pork marbling standards or pictures such as those from National Pork Producers Council (NPPC) (NPB, 2002) and is carried out by an experienced employee. The NPPC marbling standards are pictures of pork samples containing low to high intramuscular fat content with standardized numerical marbling scores from 1.0 (devoid) to 10.0 (abundant). The visual assessment however leads to inconsistencies in pork quality of different companies, increases the labor cost, has low repeatability, and is easily influenced by environment.

Therefore, more efficient and objective evaluation method for pork marbling determination would be useful for the pork industry.

In the past decade, research works related to marbling assessment of either beef or pork using machine vision approaches have been reported (Faucitano, Huff, Teuscher, Gariépy, & Wegner, 2005; Gerrard, Gao, & Tan, 1996; Liu, Ngadi, Prasher, & Gariépy, 2012; Qiao, Ngadi, Wang, Gariépy, & Prasher, 2007; Shiranita, Hayashi, Otsubo, Miyajima, & Takiyama, 2000; Tan, 2004; Toraiichi et al., 2002; Yang, Albrecht, Ender, Zhao, & Wegner, 2006; Yoshikawa et al., 2000). In comparison with beef, evaluation of marbling in pork is more challenging not only due to the much more variable color of pork but also due to the paler color of the lean part and the attenuated contrast between lean and fat (Sun, 2008, p. 126). In order to enhance the color contrast between pork marbling and lean, a chemical pre-treatment was used in Faucitano et al. (2005) so that computer image analysis (CIA) was able to be conducted on the digital color pictures (i.e. Red-Green-Blue images) of the pre-processed pork samples. A group of CIA marbling variables was calculated and some of them (such as proportion of marbling fleck area (%), the number of marbling flecks/cm<sup>2</sup>, the number of

\* Corresponding author. Tel.: +1 514 3987779.

E-mail address: [michael.ngadi@mcgill.ca](mailto:michael.ngadi@mcgill.ca) (M.O. Ngadi).



marbling flecks, total length of marbling flecks (cm), and total marbling fleck area (cm<sup>2</sup>) were significantly correlated with intramuscular fat (IMF) content. Hyperspectral imaging technology was first introduced in Qiao et al. (2007) to evaluate pork quality and marbling scores. In this work, a texture pattern analysis technique called Grey-Level Co-occurrence Matrix (GLCM) was applied to assess pork marbling levels at the wavelength of 661 nm where a significant contrast image was obtained. The angular second moment (ASM) was exploited as the texture index of marblings. Results showed that the sorted marbling score based on the ASM value was about 1.0 higher than the subjective marbling scores. In addition, ASM differentiated samples in the range of marbling scores less than 5.0 but could not segregate the samples with scores of 5.0 and 10.0. Liu et al. (2012) regarded pork marbling as a line pattern and applied a line pattern recognition technique called the wide line detector (WLD) (Liu, Zhang, & You, 2007) for the marbling extraction. The proportion of marbling (**PM**) was obtained using the WLD on digital color pictures of NPPC standards and formulated as indices of marbling scores by multiple linear regression models. Experimental results demonstrated that the established multiple linear models successfully differentiated the marbling levels of NPPC standards over the entire range. Although leave-one-out cross-validation was used to assess the generalization of the predictive models in practice, model validation on an independent sample set with the reference data from fresh pork samples was absent in this work.

The overall objective of this work was to investigate the applications of different pattern analysis techniques for pork marbling evaluation. The specific objectives were to collect digital RGB images of fresh pork loin samples; to extract different image features, i.e. line pattern and image texture, using the WLD and the improved GLCM, respectively; to establish prediction model based on extracted image features; and to compare the performance of the WLD- and the GLCM-based models for prediction of pork marbling scores.

## 2. Materials and methods

### 2.1. Samples and image acquisition

Fresh boneless pork loins ( $n = 53$ ) were obtained from a commercial packing plant in Quebec, Canada. A 2 cm thick chop was sliced from the mid portion of the Longissimus muscle and was exposed to air for a minimum of 15 min at 10 °C to allow blooming of the lean color. Marbling scores of the chop were assessed subjectively by two trained employees of the plant using the NPPC marbling standards (NPB, 2002) and the average value was used as the marbling score of the chop.

Following the subjective assessment of marbling, the RGB digital images of the chops were obtained using a NIKON D5000 digital camera with a predefined condition under a standardized environment. The spatial resolution of the camera was 300 dpi (dots-per-inch)  $\times$  300 dpi. Both surfaces of each chop were imaged for subsequent analysis.

### 2.2. Image preprocessing

The main task of image preprocessing was to select the region of interest (ROI), i.e. the loin eye area in this study. The loin eye area of each image was automatically selected by applying the image segmentation method presented in Liu et al. (2012). However, reflection points which were produced by the residual water on the chop surface were found in ROI and could affect the analysis results. In order to control the error caused by reflection points and keep the research as a single-factor study, the reflection points were

removed from the ROI manually. The final ROI used in this study is the loin eye area without reflection points. All operations of image processing and data analysis in this study were performed using MATLAB 7.3.0 (The MathWorks, Inc., Mass., USA).

### 2.3. Feature extraction

After image preprocessing, image features were extracted from three channels (red, green, and blue) of RGB images for pattern analysis and measurement. In this study, two different feature extraction methods, i.e. the wide line detector (WLD) (Liu et al., 2012) and the gray-level co-occurrence matrix (GLCM) (Qiao et al., 2007), were employed to detect line patterns and image texture, respectively.

#### 2.3.1. Wide line detector

The WLD segmented marblings from ROI in each channel of the RGB image by detecting lines as

$$L(x_0, y_0; r, t) = \begin{cases} g - m(x_0, y_0; r, t), & \text{if } m(x_0, y_0) < g \\ 0, & \text{otherwise} \end{cases}, \quad (1)$$

$$m(x_0, y_0; r, t) = \sum_{\substack{y_0-r \leq y \leq y_0+r \\ x_0-r \leq x \leq x_0+r}} s(x, y, x_0, y_0; r, t), \quad (2)$$

$$s(x, y, x_0, y_0; r, t) = \begin{cases} n(x, y, x_0, y_0; r), & \text{if } I_c(x_0, y_0) - I_c(x, y) \leq t \\ 0, & \text{if } I_c(x_0, y_0) - I_c(x, y) > t \end{cases}, \quad (3)$$

$$n(x, y, x_0, y_0; r) = \begin{cases} 1/\pi r^2, & \text{if } (x - x_0)^2 + (y - y_0)^2 \leq r^2 \\ 0, & \text{otherwise} \end{cases}, \quad (4)$$

where  $(x_0, y_0)$  is the center of the circular neighborhood,  $(x, y)$  is any other point within the neighborhood,  $I_c(x, y)$  is the intensity of the point  $(x, y)$  in the channel  $c$ ,  $r$  is the radius of circular mask,  $t$  is the intensity contrast threshold.  $s$  is the normalized weighting comparison based on the measure of similarity between the center point and any other point in the circular neighborhood, and  $m$  is the mass of the neighborhood center  $(x_0, y_0)$ .  $g$  is the geometric threshold and was set to 0.5 in this study. The output of the WLD was the isotropic line response  $L$  with the range of 0–0.5. A global threshold ( $l$ ) was used for image post-processing to remove objects with lower  $L$ . Therefore, three parameters involved in the WLD based feature extraction, i.e. the radius of circular mask ( $r$ ) which determines the maximum width of marblings can be detected, the intensity contrast threshold ( $t$ ) which defines the minimum visibility of marblings can be detected, and the global threshold ( $l$ ) which controls the noise level in the detected marblings.

The proportion of marblings (**PM**) was used as image feature index to measure marbling scores, which was defined as

$$\text{PM} = \frac{\text{area}(\text{marblings})}{\text{area}(\text{ROI})}, \quad (5)$$

where  $\text{area}(\text{marbling})$  is the total number of pixels of detected marbling, and  $\text{area}(\text{ROI})$  is the total number of pixels of the corresponding ROI. The principle behind the calculation of **PM** is comparable to human vision scaling marbling with respect to the lean background.

#### 2.3.2. Improved GLCM

The image texture indices of pork samples were derived from the grey-level co-occurrence matrix (GLCM) which was created

from each channel of the RGB image. GLCM is an image analysis technique which specifies image texture information by relative frequencies with two neighboring pixels separated by a distance in the calculation window (Haralick, 1979). This technique is widely used in food engineering, food quality and safety control to extract image texture characteristics of an object of interest (ElMasry, Wang, ElSayed, & Ngadi, 2007; Mateo, O'Callaghan, Gowen, & O'Donnell, 2010; Naganathan et al., 2008). In application, a sub-region with a regular shape (e.g. a rectangle area) is normally selected from the object of interest (such as meat, vegetables and fruits) as the ROI by hand to conduct GLCM texture analysis. However, such manual ROI selection could introduce error to the analysis and result in a biased conclusion.

In order to eliminate the influence of manual ROI selection, an improved GLCM was used in this study to extract image texture features from the irregular-shaped ROI (as shown in Fig. 2). The improved GLCM,  $\mathbf{P}_{D,\theta}$ , was defined as a square matrix with elements  $(i,j)$  specifying how often the 2 Gy levels  $i$  and  $j$  occur between two non-background pixels separated by a distance  $D$  along a given direction  $\theta$  over an image. A total of 8 Gy levels were used to scale the grey-scale values in each channel of the pork sample image. The image texture information was evaluated by different measurement of GLCM, i.e.

$$\text{Contrast} = \sum_{ij} (i-j)^2 \mathbf{P}_{D,\theta}(i,j), \quad (6)$$

$$\text{Correlation} = \frac{\sum_{ij} (ij) \mathbf{P}_{D,\theta}(i,j) - \mu_i \mu_j}{\sigma_i \sigma_j}, \quad (7)$$

$$\text{ASM} = \sum_{ij} [\mathbf{P}_{D,\theta}(i,j)]^2 \quad (8)$$

and

$$\text{Homogeneity} = \sum_{ij} \frac{\mathbf{P}_{D,\theta}(i,j)}{1 + (i-j)^2}, \quad (9)$$

where  $(\mu_i, \mu_j)$  and  $(\sigma_i, \sigma_j)$  are the means and standard deviations of the value of element  $(i,j)$  in the row and column direction in the improved GLCM, respectively. In the GLCM-based feature extraction, the measurement of GLCM obtained by Eqs. (6–9) were used as image texture feature indices (**GI**) to estimate the marbling scores of pork samples.

#### 2.4. Modeling and evaluation of models

The relationship between marbling scores and image feature indices (**PM** or **GI**) of pork samples was calculated over three channels by Pearson's correlation coefficients. All pork samples ( $n = 53$ ) were divided into two sets, the calibration set (36 samples) and the validation set (17 samples). The multiple linear regression (MLR) algorithm was employed on the calibration set to establish the prediction models for pork marbling scores based on different image feature indices derived from each channel of pork sample images. The MLR model was defined as

$$\hat{Y} = a_0 + \sum_{c=r,g,b} a_c \mathbf{X}_c, \quad (10)$$

where  $\hat{Y}$  is the vector of predicted marbling scores,  $\mathbf{X}_c$  ( $c = r,g,b$ ) is the vector of image feature indices (**PM** or **GI**) of the calibration or validation set at the channel  $C$ ,  $a_0$  is the constant term and  $a_c$  is the regression coefficient on the channel  $C$ .

The predictive ability of the MLR models was evaluated by four parameters, i.e. the root mean square error of calibration (RMSEC), the root mean square error of validation (RMSEV), the correlation coefficient between the predicted and measured marbling scores of the calibration ( $R_c$ ) and validation ( $R_v$ ) set. A good model should have low RMSEC and RMSEV, high  $R_c$  and  $R_v$ , and a small difference between RMSEC and RMSEV.

### 3. Results and discussion

#### 3.1. Image of pork sample

A total of 53 pork loin chops were used in this study. Two marbling scores of each pork sample were obtained from two experienced graders and the average value was used as the measured marbling score. Table 1 shows the statistics of measured marbling scores for all pork samples. All pork loin chops were imaged in RGB format on both sides and the ROI of each image was segmented in image preprocessing. Fig. 1 shows an example of the RGB images of a pork loin chop on both sides. The segmented ROI of one side (Fig. 1(a)) is shown in a color image in Fig. 2, as well as in grey-level images at the red, green, and blue channels, respectively. In grey-level images, both green and blue channels produced a better contrast between lean and marbling fat than the red channel.

#### 3.2. Proportion of marblings based on WLD

The wide line detector defined by Eqs. (1–4) was applied to extract marblings from the ROI at red, green, and blue channels of the digital color image of pork sample. Fig. 3 shows the marbling detection results of the RGB images at each channel using the WLD where the intensity contrast threshold  $t$  is 4, the radius of circular mask  $r$  is 30, and the global threshold  $l$  is 0.3. The parameters of the WLD were determined based on trial and error and the same values of parameters were used for all channels.

Proportion of marblings (**PM**) was calculated for all pork samples according to Eq. (5) and the results were listed in Table 2. Pearson's correlation coefficients between **PM** and marbling scores

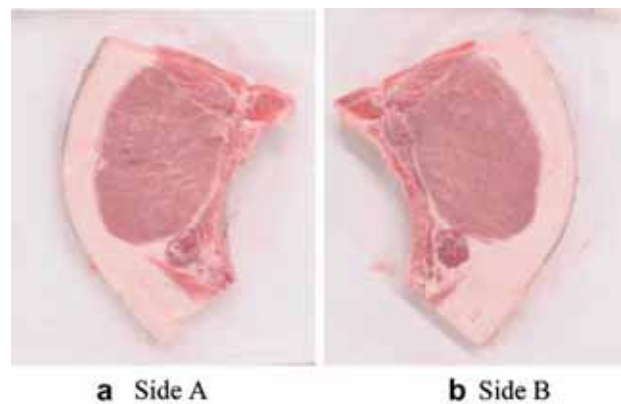
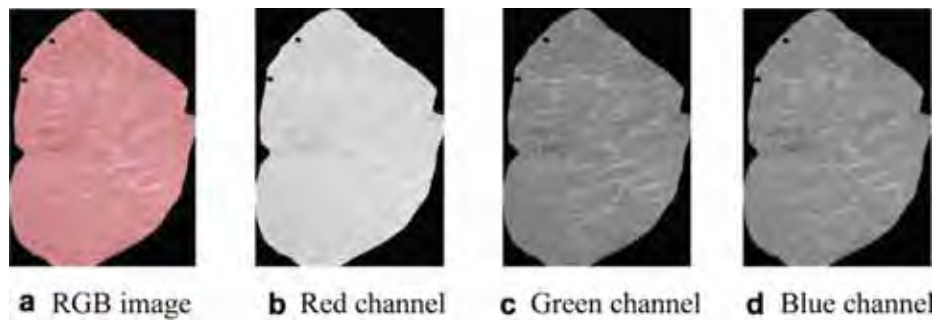


Fig. 1. Images of pork loin chops at two sides.

Table 1  
Statistics of measured marbling scores.

Parameter	Maximum	Minimum	Mean	Standard deviation
Marbling score	5.25	1.00	2.60	0.90



**Fig. 2.** ROI (marked as non-black pixels) of pork sample at different channels. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of all the samples were 0.90, 0.93, and 0.92 at the red, green, and blue channels, respectively. The obtained high correlations ( $R > 0.9$ ,  $p << 0.0001$ ) showed the strong relationship between **PM** and marbling scores. This was consistent with the result obtained by Faucitano et al. (2005) where the proportion of marbling fleck area (%) was significantly correlated with IMF content ( $R = 0.60$ ,  $p < 0.001$ ). The much higher correlation coefficients obtained in this study indicated the capability of the WLD-based pattern recognition technique for the prediction of pork marbling scores.

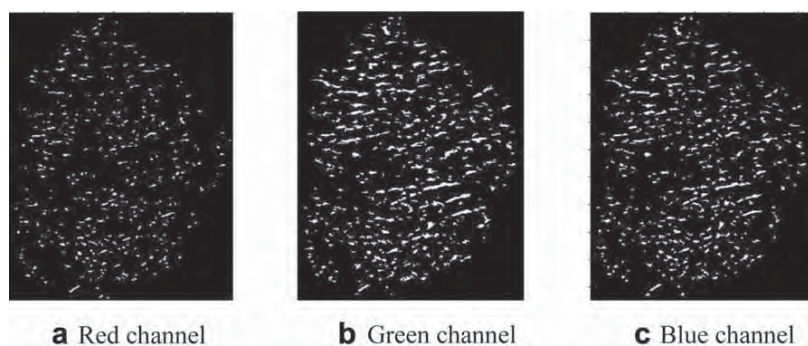
Three linear regression models, i.e. WLRr, WLRg, and WLRb, were developed using the **PM** values at the red, green, and blue channels, respectively, while one multiple linear regression model WMLR developed based on all three channels' **PM**. The regression coefficients of each model are listed in Table 3 and the corresponding prediction results shown in Table 4. The high correlation coefficients ( $R_c$  and  $R_v$ ) and low RMSEC and RMSEV from all established models showed the strong explanatory and predictive capability of **PM** for determination of pork marbling scores, and thereby indicated the effectiveness of the line pattern features for pork marbling prediction. The WLRg model established at the green channel outperformed the models at the red (WLRr) and the blue (WLRb) channels on both calibration and validation sets, which was consistent to the correlation results between **PM** and marbling scores of all pork samples. The WMLR model based on all three channels produced the same  $R_c$  as the WLRg model (0.94) and a slightly lower  $R_v$  than the WLRg model (0.93 vs. 0.94). Fig. 4 shows the plots of measured and predicted marbling scores from **PM** using the two models. The very high similarity between the two plots indicated that the **PM** obtained from the green channel had the most explanatory power, and the multiple linear regression model WMLR could be replaced by the simple linear regression model at the green channel (WLRg).

### 3.3. Image texture features based on the improved GLCM

The improved GLCM-based image texture feature extraction method was applied on the ROI at all three channels of the pork image to obtain the image texture feature index (**GI**) of pork sample. Four directions ( $\theta = [0^\circ, 45^\circ, 90^\circ, 135^\circ]$ ) and 13 steps ( $D = [1, 5:5:60]$ ) were used to calculate  $P_{D,\theta}$  and four measurements (*Contrast*, *Correlation*, *ASM*, and *Homogeneity*) were evaluated for each  $P_{D,\theta}$ .

Pearson's correlation coefficients between **GI** and marbling scores of all samples were calculated for all parameter combinations at three channels. The best correlation coefficients along each direction are shown in Table 5. The green channel produced the highest correlation coefficient in each direction and the overall best correlation coefficient between **GI** and marbling score ( $R = 0.69$ ) was obtained in the horizontal direction ( $\theta = 0^\circ$ ) with the distance of 25 pixels when the measurement of *Contrast* was used. The statistics of the corresponding **GI** (denoted as **GI**<sub>0</sub>) were listed in Table 6.

Linear regression models based on the **GI**<sub>0</sub> listed in Table 6 were developed at the red (GLRr), green (GLRg), and blue (GLRb) channels as well as the combined RGB channels (GMLR). The regression coefficients of each model are listed in Table 7 and the prediction results of the four GLCM based linear regression models are shown in Table 8. The multiple linear model based on the combined RGB channels (GMLR) outperformed the three simple linear models (GLRr, GLRg, and GLRb). Among the three simple models, the model at the green channel (GLRg) produced the best prediction results which had the same  $R_v$  as the GMLR model (0.79) and a slightly lower  $R_c$  than the GMLR model (0.72 vs. 0.73). Plots of measured and predicted marbling scores from **GI**<sub>0</sub> using the models GMLR and GLRg are shown in Fig. 5. A similarity between the two plots was observed, which showed that the **GI**<sub>0</sub> obtained from the green channel was most explanatory and the simple linear model at the



**Fig. 3.** Extracted marbling of pork at different channel.

**Table 2**  
Statistics of PM for all samples and different sets.

Parameter	Set	Channel	Maximum	Minimum	Mean	Standard deviation	
PM	All samples	Red	0.055	0.008	0.022	0.010	
		Green	0.096	0.029	0.055	0.016	
		Blue	0.095	0.032	0.056	0.015	
	Calibration	RGB	0.096	0.008	0.044	0.021	
		Red	0.055	0.007	0.022	0.011	
		Green	0.096	0.030	0.054	0.016	
	Validation	Blue	0.095	0.033	0.056	0.015	
		RGB	0.082	0.023	0.044	0.021	
		Red	0.044	0.009	0.023	0.011	
		Green		0.087	0.029	0.055	0.016
			Blue	0.087	0.032	0.056	0.015
			RGB	0.073	0.023	0.044	0.021

**Table 3**  
Regression coefficients of the WLD based LR and MLR models.

Channel	Model	$a_0$	$a_r$	$a_g$	$a_b$
Red	WLRr	0.84	80.95	0	0
Green	WLRg	-0.41	0	56.32	0
Blue	WLRb	-0.68	0	0	59.46
RGB	WMLR	-0.62	-17.63	70.40	-2.91

**Table 4**  
Prediction results of the WLD based LR and MLR models.  $p < 0.0001$  for each  $R$  value.

Model	$R_c$	RMSEC	$R_v$	RMSEV
WLRr	0.90	0.42	0.94	0.33
WLRg	0.94	0.33	0.94	0.36
WLRb	0.93	0.34	0.92	0.37
WMLR	0.94	0.33	0.93	0.37

green channel could replace the multiple linear model based on the combined RGB channels.

### 3.4. Discussion

Both the wide line detector and the gray-level co-occurrence matrix are pattern analysis techniques. In this study, the WLD extracted marbling as line patterns with position and thickness information, while the GLCM measured marbling as the texture of the pork image. Comparing the prediction results based on the two

**Table 5**  
Best correlation coefficients ( $[R;G;B]$ ) between marbling score and GI along different directions.

Direction	Measurement	Step	$R ([R;G;B])$
0°	Contrast	25	[0.53;0.69;0.56]
45°	Contrast	15	[0.50;0.66;0.55]
90°	Homogeneity	10	[-0.49;-0.60;-0.55]
135°	Homogeneity	10	[-0.48;-0.59;-0.54]

**Table 6**  
Statistics of GI<sub>0</sub> for all samples and different sets.

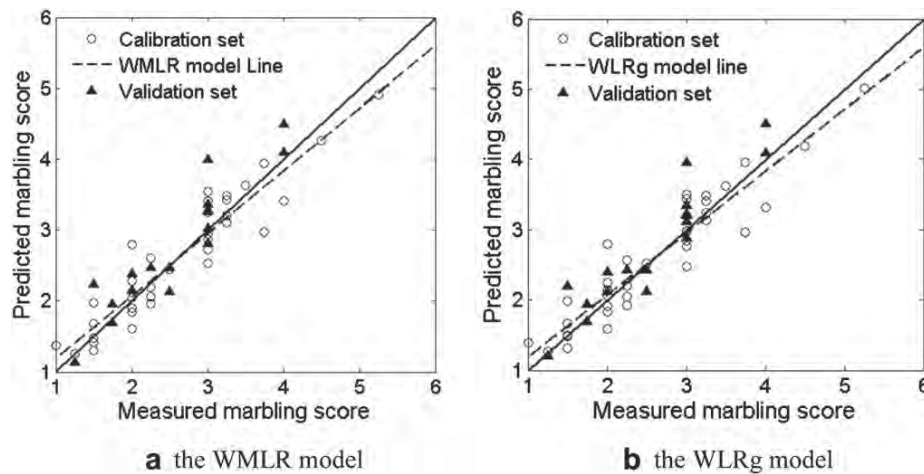
Measurement	Set	Channel	Maximum	Minimum	Mean	Standard deviation	
Contrast	All samples	Red	1.924	0.869	1.222	0.197	
		Green	0.943	0.510	0.661	0.105	
		Blue	0.977	0.432	0.649	0.105	
	Calibration	RGB	1.281	0.604	0.844	0.303	
		Red	1.640	0.890	1.207	0.169	
		Green	0.867	0.510	0.648	0.087	
	Validation	Blue	0.963	0.432	0.643	0.102	
		RGB	1.157	0.611	0.833	0.293	
		Red	1.924	0.869	1.255	0.248	
		Green		0.943	0.521	0.689	0.134
			Blue	0.977	0.469	0.661	0.115
			RGB	1.281	0.620	0.868	0.326

**Table 7**  
Regression coefficients of the GLCM based LR and MLR models.

Channel	Model	$a_0$	$a_r$	$a_g$	$a_b$
Red	GLRr	-1.24	3.06	0	0
Green	GLRg	-2.91	0	8.28	0
Blue	GLRb	-0.93	0	0	5.27
RGB	GMLR	-2.77	-0.61	9.61	-0.41

**Table 8**  
Prediction results of the GLCM based LR and MLR models.  $p < 0.01$  for each  $R$  value.

Model	$R_c$	RMSEC	$R_v$	RMSEV
GLRr	0.52	0.84	0.69	0.61
GLRg	0.72	0.68	0.79	0.71
GLRb	0.54	0.83	0.70	0.56
GMLR	0.73	0.68	0.79	0.72



**Fig. 4.** The measured and predicted marbling scores of calibration and validation sets based on WLD.



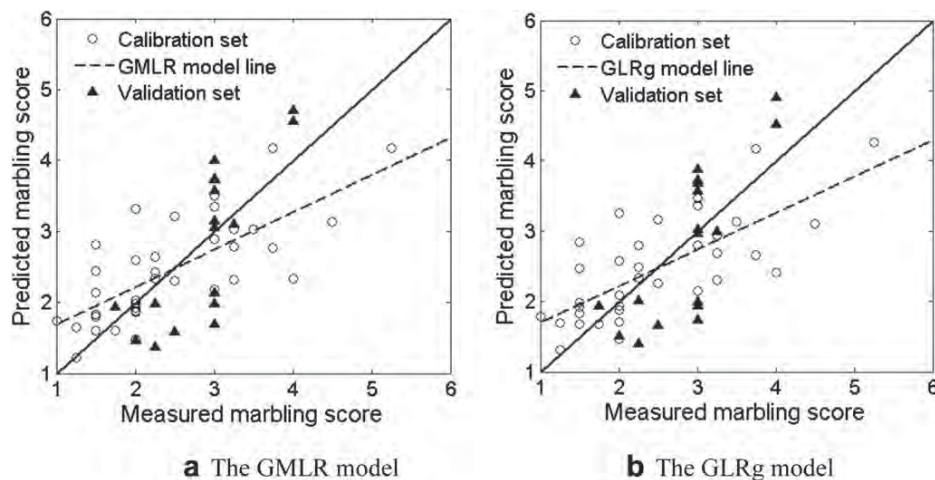


Fig. 5. The measured and predicted marbling scores of calibration and validation sets based on the improved GLCM.

techniques as shown in Tables 4 and 8, the WLD produced much better prediction than the GLCM on both calibration and validation sets for all simple linear and multiple linear models. One possible reason is that image texture features extracted using the improved GLCM included information not only from marblings but also from the lean of the pork, which could limit the ability of GLCM to predict pork marbling scores. This indicated that line pattern, as the feature extracted from the pork image to characterize marbling, could be more effective than the image texture for pork marbling prediction.

This study showed the promising potential of the WLD for pork marbling assessment. Regarding marbling as line pattern, the pork sample image with thick lines detected in a few positions could have the same **PM** values as the pork sample image with narrow lines detected in many positions according to Eq. (5), which would result in the same predicted marbling scores for both pork samples even though they have different measured marbling scores. This indicates that besides **PM** (the proportion of marbling area), the distribution of marblings (e.g. the number of marbling flecks/cm<sup>2</sup> which was used in Faucitano et al. (2005)) should also be taken into account for marbling score assessment.

Comparing the prediction results of three simple linear models, both the WLD and the GLCM based models created the best prediction of pork marbling score at the green channel. This indicated that the image feature extracted in the green channel had the best explanatory power no matter which pattern analysis technique used. The prediction results at the green channel are almost same as the results based on all three channels in Tables 4 and 8. This made it possible to develop a rapid and accurate system to assess pork marbling scores.

#### 4. Conclusion

In this study, the barriers produced by the low contrast between marbling and lean for objective and fast marbling assessment were overcome by application of effective pattern analysis techniques, i.e. the WLD and the GLCM. The results demonstrated the superior capacity of the WLD for objective assessment of pork marbling even though the predictive accuracy could be further improved. Both WLD and GLCM based models showed the green channel had the strongest prediction ability for pork marbling assessment, which indicated that simple linear model at the green channel, could replace the multiple linear model which was established based on all channels. This made it possible to develop an objective and accurate system to assess pork marblings in a real-time way. Further work will

focus on building proper industrial implementations for marbling evaluation and exploring the potential of predicting intramuscular fat content of pork by applying pattern analysis techniques.

#### Acknowledgement

The authors acknowledge the collaboration of Canadian Centre for Swine Improvement Inc. (CCSI) in data collection.

#### References

- ElMasry, G., Wang, N., ElSayed, A., & Ngadi, M. (2007). Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *Journal of Food Engineering*, 81(1), 98–107.
- Faucitano, L., Huff, P., Teuscher, F., Gariépy, C., & Wegner, J. (2005). Application of computer image analysis to measure pork marbling characteristics. *Meat Science*, 69(3), 537–543.
- Gerrard, D. E., Gao, X., & Tan, J. (1996). Beef marbling and color score determination by imaging processing. *Journal of Food Science*, 61(1), 145–148.
- Haralick, R. M. (1979). Statistical and structural approaches to texture. *Proceedings of IEEE*, 67(5), 786–804.
- Liu, L., Ngadi, M. O., Prasher, S. O., & Gariépy, C. (2012). Objective determination of pork marbling scores using the wide line detector. *Journal of Food Engineering*, 110(3), 497–504.
- Liu, L., Zhang, D., & You, J. (2007). Detecting wide lines using isotropic nonlinear filtering. *IEEE Transactions on Image Processing*, 16(6), 1584–1595.
- Mateo, M. J., O'Callaghan, D. J., Gowen, A. A., & O'Donnell, C. P. (2010). Evaluation of a vat wall-mounted image capture system using image processing techniques to monitor curd moisture during syneresis with temperature treatments. *Journal of Food Engineering*, 99(3), 257–262.
- Naganathan, G. K., Grimes, L. M., Subbiah, J., Calkins, C. R., Samal, A., & Meyer, G. E. (2008). Partial least squares analysis of near-infrared hyperspectral images for beef tenderness prediction. *Sensing and Instrumentation for Food Quality and Safety*, 2(3), 178–188.
- NPB (National pork board). (2002). *Pork quality standards*. IA, USA: Des Moines.
- Qiao, J., Ngadi, M. O., Wang, N., Gariépy, C., & Prasher, S. O. (2007). Pork quality and marbling level assessment using a hyperspectral imaging system. *Journal of Food Engineering*, 83, 10–16.
- Shiranita, K., Hayashi, K., Otsubo, A., Miyajima, T., & Takiyama, R. (2000). Grading meat quality by image processing. *Pattern Recognition*, 33, 97–104.
- Sun, D. W. (2008). *Computer vision technology for food quality evaluation* (1st ed.). Amsterdam, Boston: Elsevier/Academic Press.
- Tan, J. (2004). Meat quality evaluation by computer vision. *Journal of Food Engineering*, 61, 27–35.
- Toraichi, K., Kwan, P. W. H., Katagishi, K., Sugiyama, T., Wada, K., Mitsumoto, M., et al. (2002). On a fluency image coding system for beef marbling evaluation. *Pattern Recognition Letters*, 23, 1277–1291.
- Yang, X. J., Albrecht, E., Ender, K., Zhao, R. Q., & Wegner, J. (2006). Computer image analysis of intramuscular adipocytes and marbling in the longissimus muscle of cattle. *Journal of Animal Science*, 84(12), 3251–3258.
- Yoshikawa, F., Toraichi, K., Wada, K., Otsu, O., Nakai, H., Mitsumoto, M., et al. (2000). On a grading system for beef marbling. *Pattern Recognition Letters*, 21, 1037–1050.





# Rapid and non-invasive quantification of intramuscular fat content of intact pork cuts

H. Huang<sup>a</sup>, L. Liu<sup>a</sup>, M.O. Ngadi<sup>a,\*</sup>, C. Gariépy<sup>b</sup>

<sup>a</sup> Department of Bioresource Engineering, McGill University, Macdonald Campus, 21,111 Lakeshore Road, Ste-Anne-de-Bellevue, Quebec, Canada H9X 3V9

<sup>b</sup> Agriculture and Agri-food Canada, 3600 Casavant West, Saint-Hyacinthe, Quebec, Canada J2S 8E3

## ARTICLE INFO

### Article history:

Received 10 August 2013

Received in revised form

5 November 2013

Accepted 5 November 2013

Available online 12 November 2013

### Keywords:

Intramuscular fat content

Gabor filter

Gray level co-occurrence matrix

Partial least squares regression

Multiple linear regression

Distribution map

## ABSTRACT

Having acquired near infrared (NIR) hyperspectral images of intact pork loin samples through an NIR hyperspectral imaging system, the efficiency of a variety of image processing techniques including texture pattern analysis techniques were applied to process hyperspectral images so as to determine the intramuscular fat (IMF) content non-destructively. After the segmentation of region of interest (ROI), the raw spectral, texture-based spectral and textural characteristics of pork images were extracted by spectral averaging and pattern recognition techniques namely Gabor filter and improved gray level co-occurrence matrix (GLCM), respectively. First derivatives of the non-filtered and the Gabor filtered spectra were also investigated. Full waveband partial least squares regression (PLSR) was employed to determine the optimal parameters of Gabor filter and GLCM, and to select optimal wavelengths for IMF prediction. A stepwise procedure was applied to the optimal wavelengths to further optimize them to key wavelengths. Multiple linear regression (MLR) models were built based on the key wavelengths. Mean spectra and the Gabor filtered spectra outperformed GLCM. The best result, represented by correlation coefficients of calibration ( $R_c$ ), cross validation ( $R_{cv}$ ) and prediction ( $R_p$ ) of 0.89, 0.89, and 0.86, respectively, was achieved using the first derivative of Gabor filtered spectra at 1193 and 1217 nm. To visualize the IMF content in pork, the distribution maps of IMF content in pork were drawn using a mean spectra-based MLR model. These promising results highlight the great potential of NIR hyperspectral imaging for non-destructive prediction of IMF content of intact pork.

© 2013 Published by Elsevier B.V.

## 1. Introduction

Intramuscular fat (IMF) refers to the fat deposited inside a piece of muscle, so accordingly intramuscular fat content is defined as the mass of IMF, including the visible fat and the non-visible fat in the muscle cell. IMF content influences the cooking quality of pork (flavor, juiciness, etc.), eating satisfaction of consumers, and even health-related issues. Thus, different levels of IMF content can lead to different levels of consumer acceptance [1,2]. Non-invasive and rapid determination of IMF content of pork chops would allow commercial cuts to be classified, screened, and assigned to a proper retail category according to different market targets, thereby enhancing their market allocation and reducing handling

costs. At present, the assessment of IMF content of pork involves solvent-based lipid extraction, which is a time consuming, labor intensive and environmentally harmful process. Since chemical extraction-based IMF prediction is not suitable for fast and non-destructive assessment of IMF content of pork, random sampling is typically used in the evaluation of IMF content of pork products. As this leads to non-optimal grading of pork cuts, it would be beneficial to the pork industry to develop a non-destructive, real-time, rapid, and accurate method for predicting IMF content of pork.

Studies have been conducted to evaluate IMF content of pork meat through spectroscopic means [3–7]. Savenije et al. [6] applied a near infrared (NIR) reflectance spectrophotometer to the determination of IMF content of meat obtained from three contrasting breeds. The second derivative of the reflected spectra was used to build prediction models. Correlation coefficients of calibration ( $R_c$ ) between the measured and predicted IMF contents ranged from 0.70 to 0.86, while correlation coefficients of validation ( $R_v$ ) ranged from 0.63 to 0.76. Prevolnik et al. [4] and Barlocco et al. [5] found that estimate of IMF levels by NIR spectroscopy were more accurate when derived from minced pork muscle other than intact pork muscle, highlighting the limitations of spectroscopic

Abbreviations: **DMG**, the first derivative of Gabor-filtered mean spectra; **DMS**, the first derivative of non-filtered mean spectra; **GF**, Gabor filter; **Gla**, angular second moment of GLCM; **Glh**, homogeneity; **Glt**, contrast of GLCM; **Gin**, correlation of GLCM; **MG**, Gabor-filtered mean spectra; **MS**, non-filtered mean spectra; **WLD**, wide line detector.

\* Corresponding author. Tel.: +1 514 3987779; fax: +1 514 3988387.

E-mail address: [michael.ngadi@mcgill.ca](mailto:michael.ngadi@mcgill.ca) (M.O. Ngadi).

technique for determination of IMF content of pork. Given the confined detected area of the spectroscopic system, and the heterogeneous distribution of fat in pork muscle, only limited information could be obtained. Therefore, to obtain sufficient data from pork samples either destructive homogenization of samples or repetition of spectroscopy-based evaluation is required.

In the face of these limitations, the emerging technique of hyperspectral imaging was exploited to inspect pork quality [8–12]. By integrating both conventional spectroscopy techniques with imaging techniques and enlarging the detection field, hyperspectral imaging overcomes the limitations of spectroscopy, making it possible to identify the spectral details of different chemical components at specified locations in a product [13,14]. From the resultant data cube (*hypercube*), image parameters or spectra can be extracted and analyzed in order to determine the chemical attributes or physical properties of the tested object. Moreover, the hypercube can be stored for further analysis. In addition, no sample preparation is needed to apply hyperspectral imaging. Considering the great potential of this technology, there has been mounting interest in applying hyperspectral imaging to pork quality control [15].

Qiao et al. [9] introduced visible (VIS)/near-infrared (NIR) hyperspectral imaging to classify pork quality and assess pork marbling objectively. Image processing approach namely gray level co-occurrence matrix (GLCM) was adopted to grade pork samples into different meat quality categories classification results of 75–80% were achieved. Liu et al. [10] developed a Gabor filter-based hyperspectral imaging system to grade pork samples into the same meat quality categories. The work obtained a classification accuracy of  $84 \pm 1\%$ , which improved the earlier result of Qiao et al. [9] by 4%. This implied that effective pattern recognition techniques would enhance the ability of hyperspectral imaging for assessment of pork quality. Promising results in these studies again emphasized the capacity of hyperspectral imaging as the basis for prediction of pork quality attributes.

The functional bonds of C–H and O–H in fat are closely linked to some frequencies in the NIR region, accounting for why the use of NIR hyperspectral imaging for IMF content prediction might produce better results than hyperspectral imaging in the visible (VIS) region. Kobayashi et al. [16] and Wold et al. [17] successfully applied NIR hyperspectral imaging for control of fat content in beef. However, hyperspectral imaging of pork presents a greater challenge, compared to beef, due to the lower contrast between fat and lean in pork and the spectral overlaps in the NIR region exhibited by the main constituents of meat (lipid, water, and protein) [19,20]. Nonetheless, Liu et al. [18] reported the use of the hyperspectral imaging technique to predict the IMF content of pork. The reflected images in a specific wavelength interval were accumulated to enhance the contrast between the fat and the non-fat areas. A feature detection method namely the wide line detector then served to detect fat flecks of pork, yielding a prediction accuracy of 0.91 through a linear regression modeling. While this study indicated the possibility of rapid assessment of the intramuscular fat content of pork using NIR hyperspectral imaging, the authors mentioned the need for more sample measurements and wavelength optimization.

The aim of the present study was to investigate the potential of NIR hyperspectral imaging (900–1700 nm) for quantification of IMF content in pork. For this purpose, normal averaging of spectra and two texture pattern techniques namely Gabor filter and improved GLCM were investigated in terms of their ability to draw useful information from NIR spectral images. Partial least squares regression (PLSR) in the whole wave range was used to select feature wavelengths, after which a stepwise procedure was used to further reduce the dimension of optimal wavelengths. The distribution of IMF content of pork was visualized using the built optimal multiple linear regression model.

## 2. Experimental section

### 2.1. Sample collection and determination of IMF content

Pork samples from *longissimus thoracis* muscle were collected from the carcasses of 83 pigs which were raised on a local farm. No differing treatment was applied to the animals. To increase the variability in IMF content in the *longissimus* muscle, carcasses showing large differences in back fat thickness between the 3rd/4th last ribs were selected for dissection. At 24 h post-mortem, thin pork slices at the 3rd/4th last thoracic rib of the *longissimus dorsi* (LD) were collected for determination of IMF content. Concurrently, loin chops with 2–2.5 cm in thickness were dissected from the same anatomical location (the 3rd/4th last thoracic rib), individually wrapped in vacuum packs, and transported to the Hyperspectral Imaging Laboratory, McGill University, Montreal, QC, Canada for image acquisition. The shipping was conducted at 4 °C.

Peripheral fat and surrounding muscle were removed from the thin pork slices. The remaining trimmed muscles were ground. The fresh weight based IMF content of the minced fresh pork was measured using ethanol and dichloromethane-based Soxhlet extraction [21]. These IMF contents served as reference values for corresponding loin samples.

### 2.2. NIR hyperspectral imaging system

An NIR hyperspectral imaging system (Fig. 1) was used to capture NIR images of pork loins. The system consisted of an InGaAs camera (XenICs, Belgium), a spectrograph (Headwall photonics, USA, 900–1700 nm), a conveyer (Donner 2200 series, Donner Mfg. Corp., USA), a real-time control motor (MDIP22314, Intelligent motion system Inc., USA), two 50 W tungsten halogen lamps (vertically 25 cm, and horizontally 8 cm away from the samples), a supporting frame, and a computer. Incident illumination was at an angle of 45° to the sample. The distance between lens and the sample was 40 cm. The directions of the lamps were set up to make sure that the light sources are enabled in every part of the sample. Software for data acquisition and conveyer control was installed on the computer (Hyperspec, Headwall Photonics Inc., USA). The system was run at a room temperature (20–25 °C).

The line-scanning pushbroom system employed was operated in a reflectance mode. With the conveyer moving at a predefined speed, the sample was scanned line by line and the reflected light from samples was quantified at wavelengths ranging from 900 to 1700 nm by the spectrograph. A three-dimensional *hypercube* was formed by combining spatial information from the InGaAs camera and spectral information from the spectrograph. The spectral resolution for the system was 4.8 nm while the spatial resolution was roughly 0.6 mm. The spectral dimension covered 167 bands. The number of pixels covered by one spatial dimension was determined by the morphological size of the detected object. The generated *hypercube* was saved in a band-interleaved-by-line (BIL) format, from which NIR images and spectra for each pixel could easily be extracted.

### 2.3. Image acquisition and calibration

No sample preparation was needed before image acquisition. Each sample was taken out of the refrigerator (4 °C), scanned for few seconds to collect images, and put back to the refrigerator right after the image collection. Considering the distance between light source and the sample, and the speed of image acquisition, the temperature of the sample was assumed to remain nearly at 4 °C and no significant changes occurred to the sample during imaging. Prior to the collection of hyperspectral images from the 2–2.5 cm thick loin chops, dark and white reference images were captured to correct error caused by dark current of the system and

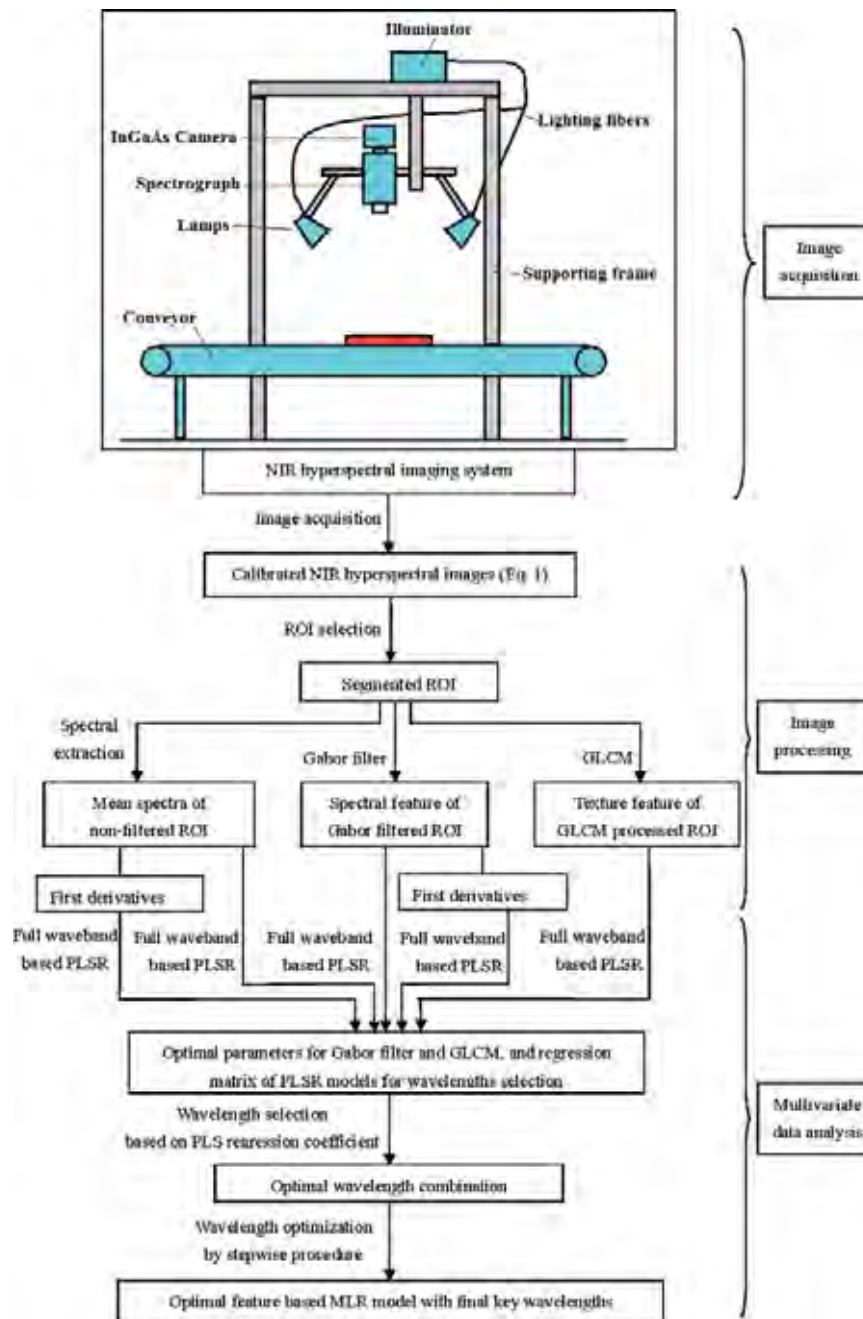


Fig. 1. Flow chart of data acquisition and analysis.

to transform the reflected light intensity to reflectance. The dark image ( $\mathbf{D}$ , with reflectance  $\sim 0\%$ ) was obtained by covering the lens with an opaque cap. The white reference ( $\mathbf{W}$ , with reflectance  $\sim 99\%$ ) was obtained by adopting a ceramic tile as photographed object. Original NIR images ( $N_0$ ) of samples were collected by placing sample on a dark board whose reflectance was close to 0%. Image acquisition of each sample took approximately 4 s. Corrected NIR images ( $\mathbf{N}$ ) were obtained by algebraic calculus (Eq. (1)) of the light intensities of each pixel  $\{i, j\}$  in dark ( $\mathbf{D}(i, j)$ ), white ( $\mathbf{W}(i, j)$ ) and original NIR images ( $N_0(i, j)$ ). The reflectance of pixels in the calibrated NIR images ranges from 0 to 1. Subsequent image and data analysis (Fig. 1) were conducted using the corrected hypercube.

$$N(i, j) = \frac{N_0(i, j) - D(i, j)}{W(i, j) - D(i, j)} \quad (1)$$

## 2.4. Image processing

The spectral averaging-based and pattern technique-assisted image processing and subsequent multivariate statistical analysis are summarized in Fig. 1. Given the excessive noise of images at 900–935 nm and 1655–1700 nm, only those images obtained over the range of 940–1650 nm were used for image analysis, starting with image preprocessing. All operations of image processing and multivariate statistical analysis in this study were performed using MATLAB 7.3.0 (The MathWorks, Inc., Mass., USA).

### 2.4.1. Image preprocessing

Regions of interest (ROI) were segmented, from which the spectral and texture features were extracted. ROI Segmentation with the aim of isolating the targeted lean and IMF portion of the

meat from other portions is an important step in image processing, as the selected ROI will affect all further analysis. The peripheral fat, surrounding muscle, shadows and other backgrounds were removed and replaced by a homogenous black background. ROI segmentation in this study was conducted according to the method presented in Liu et al. [22]. The threshold value was automatically set according to the mean and standard deviation of reflectance of each image.

2.4.2. Extraction of spectral feature

The mean spectrum of non-filtered ROI (MS) was extracted after image preprocessing. MS of each sample was a 1 × 149 (the wave range/the spectral resolution of the system+1) vector. Average spectra of the two sides of each pork sample were used as final spectral features. In general, there were over 40,000 pixels in the ROI obtained by hyperspectral imaging, hence the error introduced by subjective selection of ROI was reduced comparing to spectroscopy. To raise the ratio of signal to noise, the first derivative (Eq. (2)) was applied to the obtained non-filtered mean spectra. The resulted first derivative of mean spectra (DMS) was used for data analysis as well as MS. DMS of each pork sample was a 1 × 148 vector (the wave range/the spectral resolution of the system).

$$D_i = \frac{MS_{i+1} - MS_i}{\lambda_{i+1} - \lambda_i} \quad (2)$$

where *i* represents the number of wavelengths, *i* = 1, 2, 3, ... 148. *D<sub>i</sub>* is the first differential of the mean spectrum at *i*th wavelength. *λ<sub>i</sub>* is the *i*th wavelength, with an interval of 4.8 nm between *λ<sub>i</sub>* and *λ<sub>i+1</sub>*. *MS<sub>i+1</sub>* and *MS<sub>i</sub>* are the value of spectral response at the (*i*+1)th and *i*th wavelength, respectively.

2.4.3. Extraction of texture spectrum by Gabor filter

Proposed as an analogue of human vision [10,23], texture pattern technique Gabor filter served to extract texture features from NIR images. The two-dimensional Gabor filter is a transformation of elliptic Gaussian and sinusoidal waves, which are applied on all receptive fields in the image. Gabor filter is capable of extracting the important spatial characteristics including spatial localization and spatial frequency from images. Considering IMF as spatial characteristics in pork images, an isotropic Gabor filter (GF1, Eq. (3)) and an oriented Gabor filter (GF2, Eq. (4)) were used to process pork images at wavelengths ranging from 940 to 1650 nm [24].

$$GF1(x, y, f, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{x^2 + y^2}{2\sigma^2}\right\} \cos [2\pi f(\sqrt{x^2 + y^2})] \quad (3)$$

$$GF2(x, y, f, \sigma, \theta) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{x^2 + y^2}{2\sigma^2}\right\} \cos [2\pi f(x \cos \theta + y \sin \theta)] \quad (4)$$

where (*x*, *y*) are the coordinates of a specified pixel in a given NIR image, *f* represents the frequency of the sinusoidal wave, *σ* is the standard deviation of the Gaussian function, and *θ* is a vector which controls the orientation of the filter (*θ* = 0°, 45°, 90°, 135°). Since GF1 is an isotropic function, *θ* was not considered as a parameter for GF1. After filtering, the mean spectral response of the Gabor filtered ROI was obtained. To reduce the influence of heterogeneity in the pork sample, the average spectrum from two surfaces of each sample was obtained and used for subsequent analysis. Mean spectra from GF1 and GF2 processed images were denoted as MG1 and MG2, respectively. The first derivatives of MG1 and MG2 were calculated according to Eq. (2) and denoted as DMG1 and DMG2, respectively. DMG stands for the first derivatives of both types of Gabor filtered spectra.

2.4.4. Extraction of texture features by improved GLCM

Another texture pattern analysis technique termed gray-level co-occurrence matrix (GLCM) was investigated for processing of hyperspectral images. GLCM provides information about how often the pixel intensities occur between two pixels that are distributed by a specific distance and direction. Features derived by GLCM have been widely used in texture analysis for food quality and safety control [25–27]. In applying GLCM, the subjective selection of a regular shape (e.g. a circle area) is usually used to select the appropriate ROI [9,26]. However, as bias would be introduced by manual interference, especially for heterogeneous objects such as meat products. In the present, an advanced GLCM was applied to the full irregularly shaped ROI of pork loin to avoid the effect of subjective ROI selection.

To calculate texture features, a multi-scale GLCM matrix was derived from each image. For instance, Fig. 2(a) illustrates how an 8-level GLCM matrix was calculated from a 6 × 5 matrix (i.e. 6 × 5 image). As the shadow elements shown in Fig. 2(a), with an offset direction of east (0°) and an offset distance of 1.0 between two pixels, two instances of two neighboring pixels having values of 1 and 3 occurred. Hence, the element {1, 3} in the GLCM contains the value 2. The size of the generated GLCM matrix depends on the number of gray scales considered. Usually, eight scales were considered, resulting in an 8 × 8 square GLCM matrix.

Eight scale levels, four offset directions (Fig. 2(b), 0°, 45°, 90°, 135°), and 11 offset distances between 1 and 3–30 pixels, with a stepsize of three pixels were tried in generating the GLCM matrices. The minimum and the maximum intensities of the input image delimited the scale, which was divided into eight equal levels. For each image, 52 (4 directions × 13 steps = 52) different matrices were generated. The image texture features (GI) included four measurements calculated from obtained GLCM matrix (Eqs. (5)–(8)): contrast (GI<sub>c</sub>), correlation (GI<sub>n</sub>), angular second moment (ASM, GI<sub>a</sub>), and homogeneity (GI<sub>h</sub>), respectively.

$$Contrast = \sum_{ij} (i-j)^2 M_{D,\theta}(i,j) \quad (5)$$

$$Correlation = \frac{\sum_{ij} (ij) M_{D,\theta}(i,j) - \mu_i \mu_j}{\sigma_i \sigma_j} \quad (6)$$

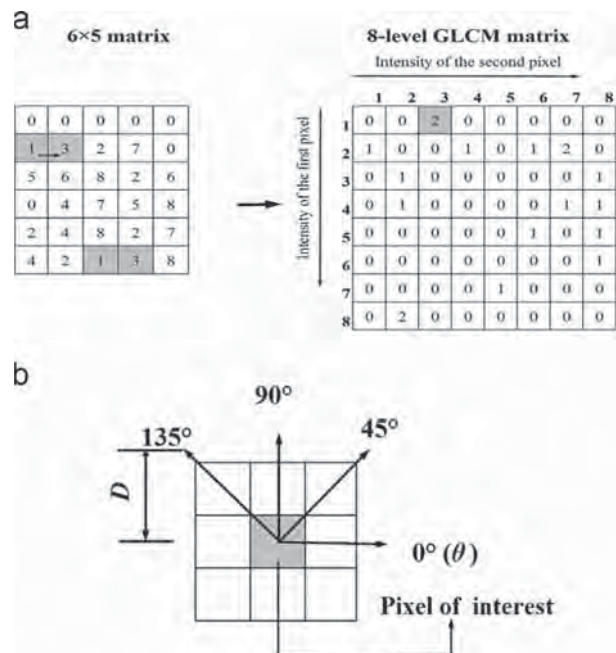


Fig. 2. Principle of GLCM matrix. (a) Form of 8-level GLCM matrix and (b) distribution of pixel pair.



$$ASM = \sum_{ij} [M_{D,\theta}(i,j)]^2 \tag{7}$$

$$Homogeneity = \sum_{ij} \frac{M_{D,\theta}(i,j)}{1+(i-j)^2} \tag{8}$$

where  $(i, j)$  is the coordination of the pixel of interest,  $D$  is the given distance between two pixels,  $\theta$  is the offset direction of the pair of pixels over an image ( $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ ),  $M_{D,\theta}$  is the obtained GLCM matrix,  $M_{D,\theta}(i, j)$  is the value in element  $\{i, j\}$  of  $M_{D,\theta}$ ,  $(\mu_i, \mu_j)$  are the means of  $M_{D,\theta}(i, j)$  in the row and column directions, and  $(\sigma_i, \sigma_j)$  are the standard deviations of  $M_{D,\theta}(i, j)$  in the row and column directions. As a result, 208 ( $52 \times 4 = 208$ ) plots with different parameters ( $\theta, D$ , measurement) were developed from each image. The four measurements of GLCM each served as a type of image texture feature index in the estimation of IMF content of pork.

### 2.5. Multivariate data analysis

Partial linear square regression (PLSR) and a stepwise procedure were employed in building calibration models, which served in

determining optimal parameters for GF and GLCM and key wavelengths for modeling. PLSR method is widely used to reduce the dimensionality of predictor variables and random noise, and has proved to be efficient [7]. In practice, the essential step of PLSR analysis is selecting the number of the main PLS principal components (PLS-PC) which explain the maximum fundamental relations between predict and response variables. After quantification of PLS-PC, the PLSR model is built. The parameters of GF ( $f, \sigma, \theta$ ) and GLCM ( $D, \theta$ , measurement), which provided best prediction results in PLSR models, were used as the optimal parameters. The corresponding regression coefficient was used to select optimal wavelengths.

In this study, features at all wavebands (940–1650 nm) were used for the PLSR analysis. The total 83 samples were divided into two sets, including calibration sets (56 samples) and prediction sets (27 samples). PLSR models with different features were built using calibration sets. Numbers of PLS-PC were determined when the root mean squared error of calibration (RMSEC) reached the minimum value. The robustness of the calibration models generated was tested by leave-one-out cross validation. Considering the predictive ability of models for real samples as well, data in the prediction sets were input

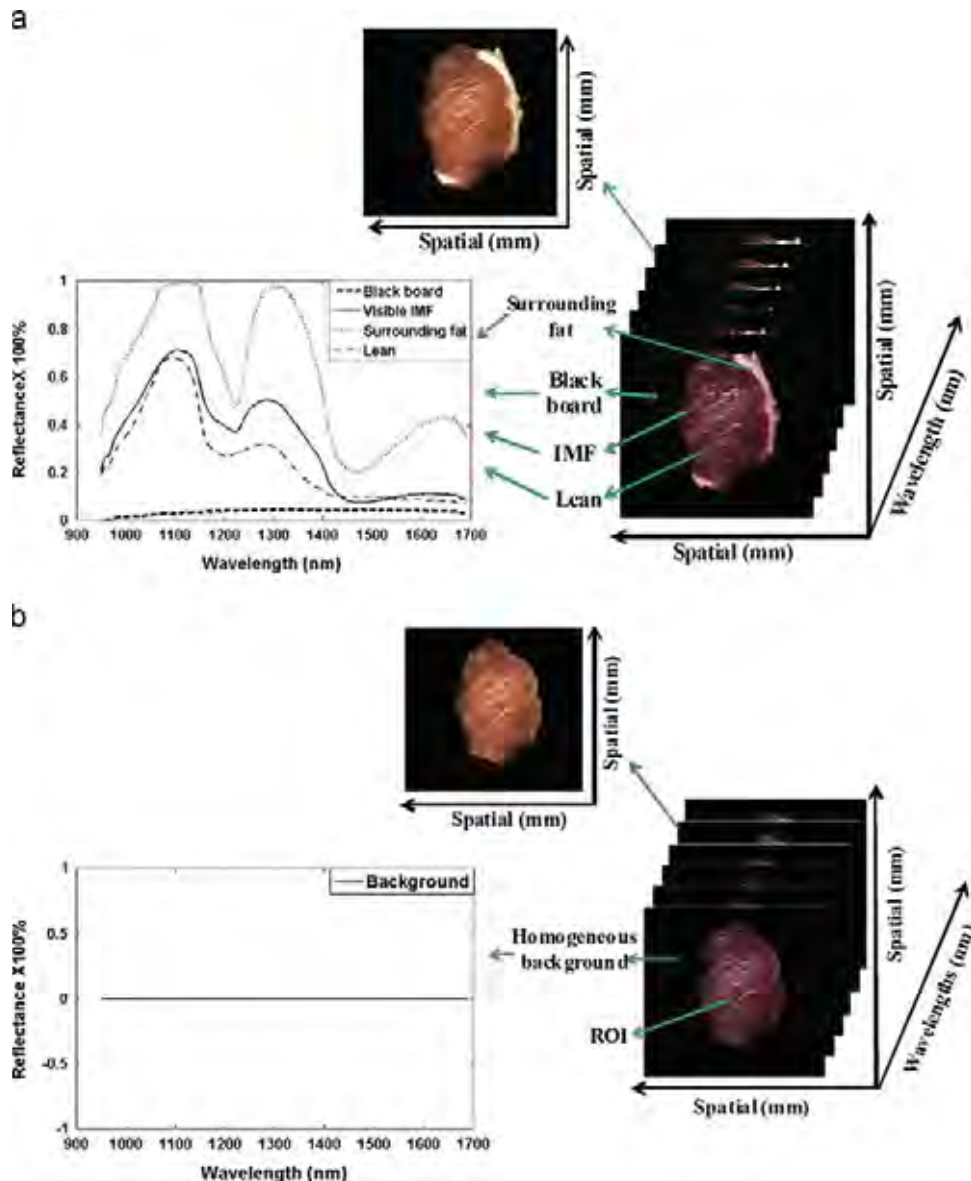


Fig. 3. Construction of hypercube and segmentation of ROI. (a) Raw hypercube, hyperspectral images, and spectra of black board, visible IMF, surrounding fat, and lean, (b) ROI and spectrum of homogeneous background pixel.



into calibration models for testing. After the numbers of PLS-PC were determined, values of GF-based or GLCM-based models that produced best results were selected along with corresponding optimal parameters of feature extraction. The efficiency of PLSR models was assessed according to following statistical values of calibration sets, cross validation sets, and prediction sets: the correlation coefficient of calibration ( $R_c$ ), cross validation ( $R_{cv}$ ), and prediction ( $R_p$ ), the root mean square error of calibration (RMSEC), cross validation (RMSECV), and prediction (RMSEP). Models with greater  $R_c$ ,  $R_{cv}$ , and  $R_p$  values, and lower RMSEC, RMSECV, and RMSEP values were preferred.

The most valuable wavelengths for each feature were selected from peaks in their individual plot of regression coefficients.

A stepwise procedure was employed to further narrow down the key wavelengths. The variables at selected key wavelengths were input into an MLR model (Eq. (9)), and their performance in calibration, cross validation, and prediction was compared. The most practical model was selected for prediction of IMF content in intact pork.

$$Y_m = b_0 + \sum_{i=1}^n X_i \times b_i \quad (9)$$

where  $b_0$  and  $b_i$  are regression coefficients,  $Y_m$  is the measured IMF content of pork samples,  $X_i$  is the variable at the  $i$ th wavelength,  $n$  represents the number of key wavelengths used, i.e. variables. The

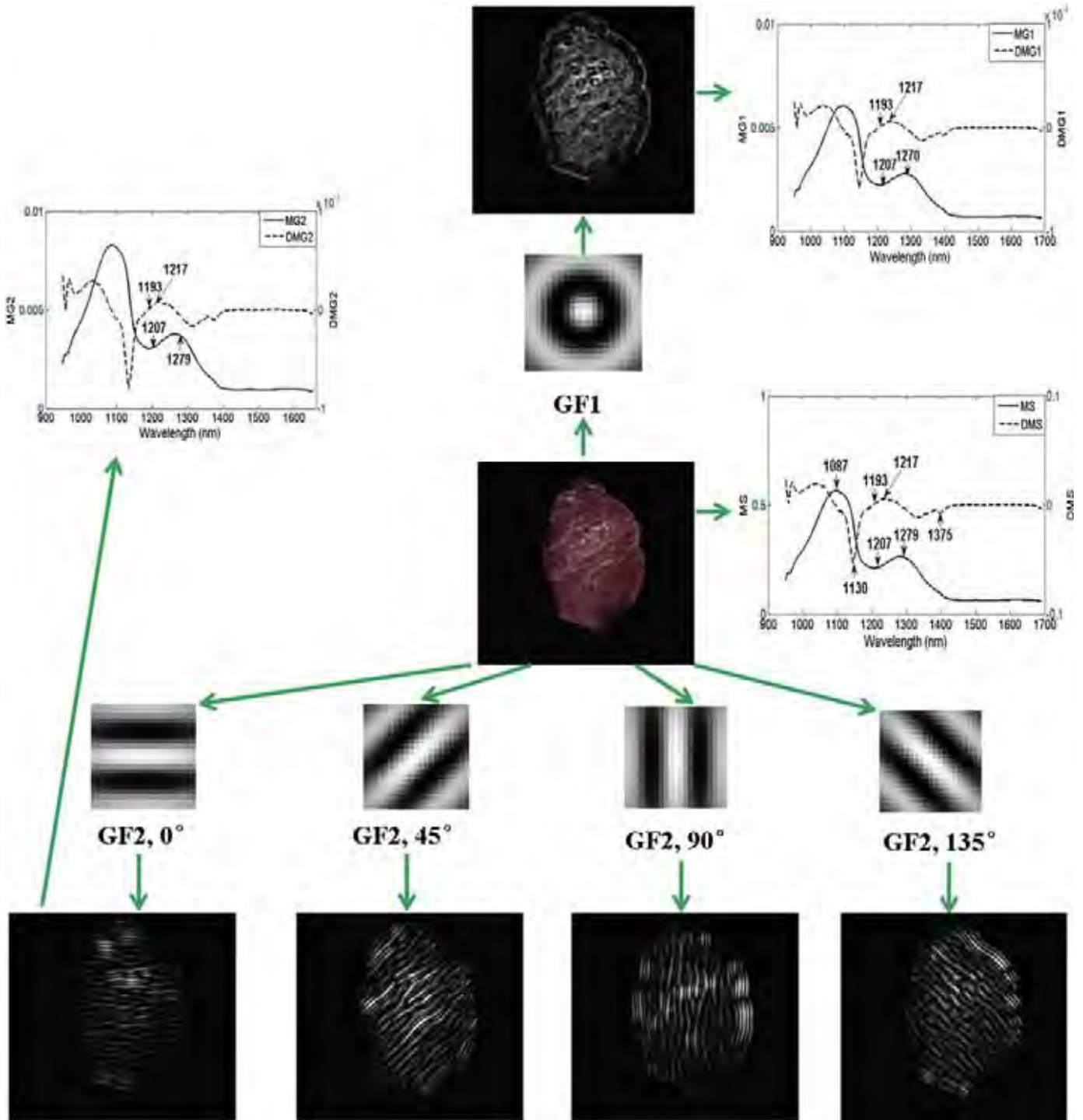


Fig. 4. The raw ROI, GF1, GF2, corresponding Gabor filtered ROI, and generated typical spectral responses.

multiple linear regression model was assessed by statistical values of  $R_c$ ,  $R_{cv}$ ,  $R_p$ , RMSEC, RMSECV, and RMSEP. Models with greater  $R_c$ ,  $R_{cv}$ , and  $R_p$  values, lower RMSEC, RMSECV, and RMSEP values, and smaller differences between RMSEC, RMSECV, and RMSEP were preferred.

2.6. Visualization of IMF content in pork by distribution map

Since pork is a heterogeneous material, the visualization of IMF distribution would allow a better interpretation of the IMF content in pork. One advantage of hyperspectral imaging is that it provides spectral information of each pixel in pork image, which means the IMF content of each pixel can be predicted by inputting the spectrum into a prediction model. Furthermore, the spectral information from hyperspectral imaging includes both external information regarding the objects of interest, as well as internal information, making hyperspectral imaging more suitable for prediction of IMF content, as IMF is distributed not just on the surface of pork, but throughout the whole pork sample. Hence, application of hyperspectral imaging has the potential to assist in the visualization of IMF content in intact pork cuts. Towards this purpose, reference data of IMF content of each pixel would be required to build an accurate prediction model. However, it is not possible to measure the IMF content of a pixel. To overcome this challenge, the best and simplest MLR model that was constructed based on mean spectrum or image parameters in previous step was selected. The MLR model was applied to the adopted spectral or image features of each pixel to predict the IMF content of each pixel in the pork image. By showing the pixel-based IMF content, the distribution map of IMF content in pork was then generated. This would help the understanding of IMF distribution in pork and also assist in conducting further detailed study at the pixel level.

3. Results and discussion

3.1. IMF content and ROI of hyperspectral images

A total of 83 pork loin chops were investigated in this study. Wide variations in IMF contents were observed for the total pork cutlets ( $n=83$ ), calibration set ( $n=56$ ), and prediction dataset

( $n=27$ ), with range of 0.51–5.8, 0.51–5.8, and 0.58–3.62, respectively. The range of prediction was covered by calibration set and the standard deviation (STD) of all three datasets was in the same level, ensuring a stable and reliable calibration model.

The conformation of *hypercube* (940–1650 nm) generated by the hyperspectral imaging system and corresponding ROI segmentation are illustrated in Fig. 3(a). A typical NIR image and the spectra of a surrounding fat pixel, an IMF pixel, a lean pixel, and a black board pixel were extracted from a *hypercube*. The reflected spectra of fat and lean showed different scales of intensity but similar features: reflected peak around 1100, 1300, and 1650 nm, and valley around 1250 and 1450 nm. The main constituents of fat and lean include lipid and water, whose peak absorption bands show a great deal of overlapping [28]. Besides, each hyperspectral image was composed of not only the external information of sample, but also the internal information acquired by penetration of NIR spectra into the sample. Since pork is a heterogeneous object, the spectrum from one pixel may contain information of lean and fat in different layers simultaneously. Pork was cut across the grain of muscle cells. Each layer included information of numerous muscle cells. This would explain the similarity in spectral characteristics between lean and fat. The reflectance of a black board pixel was near but not absolutely zero. To screen the board and surrounding fat from the image being analyzed, automatically segmented masks were applied to each NIR image across the waveband of 940–1650 nm, such that reflected values for non-ROI areas were all set to zero. The resultant spectrum of background pixel is absolutely zero in Fig. 3(b).

3.2. Spectra from raw ROI and Gabor filtered ROI and the first derivative of spectra

The raw ROI and the corresponding Gabor filtered ROI ( $f= 0.1$ ,  $\sigma= 10$ ), and typical typical mean spectra of raw ROI (**MS**) and Gabor filtered ROI (**MG1** and **MG2**), and the corresponding responses of first derivative of spectra (**DMS**, **DMG1** and **DMG2** ( $\theta= 0^\circ$ )) are illustrated in Fig. 4. ROI in each image was either filtered through the isotropic GF1 or the four orientation GF2 ( $\theta=$

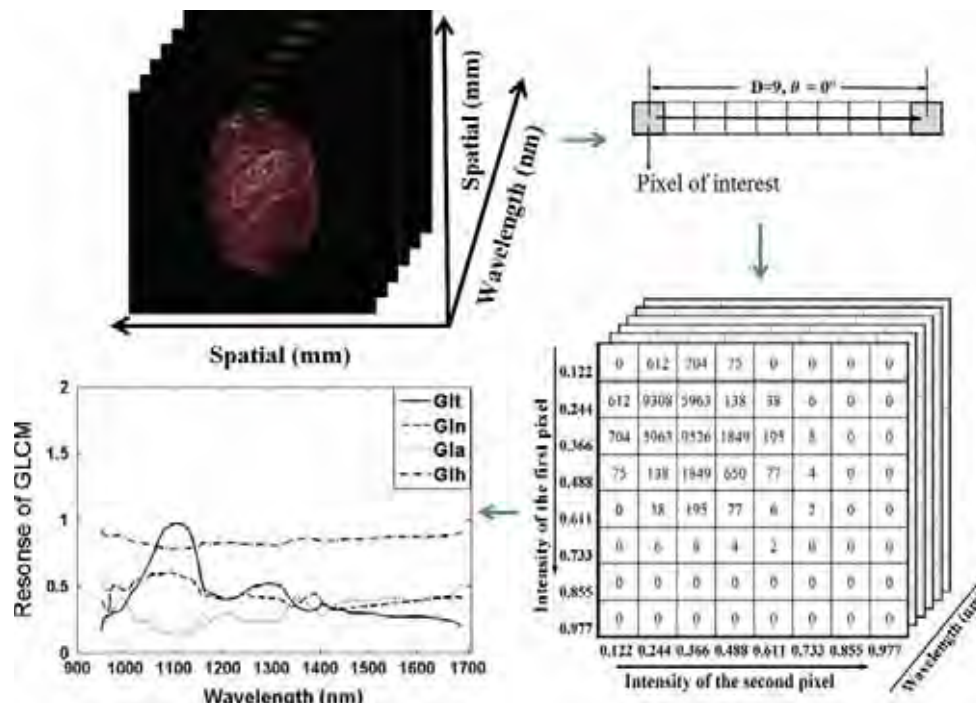


Fig. 5. Formation of GLCM matrix and corresponding texture curves.

**Table 1**  
Results of full waveband-based PLSR models using spectral and texture features.

Features	Parameters			Number of PC	Calibration		Cross validation		Prediction	
	$\sigma$	$f$	$\theta$		$R_c$	RMSEC	$R_{cv}$	RMSECV	$R_p$	RMSEP
<b>MS</b>				7	0.90	0.44	0.84	0.55	0.81	0.52
<b>DMS</b>				7	0.90	0.44	0.84	0.55	0.82	0.52
<b>MG1</b>	10	0.1		8	0.90	0.43	0.83	0.56	0.82	0.51
<b>DMG1</b>	10	0.1		7	0.90	0.44	0.83	0.56	0.82	0.52
<b>MG2</b>	10	0.1	0°	7	0.90	0.44	0.85	0.55	0.83	0.52
<b>DMG2</b>	10	0.1	0°	7	0.90	0.44	0.84	0.56	0.82	0.53
	Measurement	$D$	$\theta$							
<b>GI</b>	Contrast	9	0°	3	0.80	0.65	0.76	0.70	0.79	0.75

0°, 45°, 90°, 135°). The mean spectra of raw ROI and filtered ROI showed a similar pattern but different magnitudes. All mean spectra showed similar features: a steep peak around 1087 nm, another peak around 1279 nm, and a valley around 1207 nm. The response of first derivative of mean spectra showed similar features: a gradual peak around 1217 nm, and a steep valley around 1130 nm. **MS**, **DMS**, **MG1**, **MG2**, **DMG1**, and **DMG2** spectra of samples were used for multivariate data analysis.

### 3.3. Texture curve by GLCM

A typical 8-level GLCM matrix of a ROI was formed and the GLCM derived measurements of *contrast*, *correlation*, *ASM*, and *homogeneity* (**Git**, **Gln**, **Gla**, and **Glh**) were obtained as illustrated in Fig. 5, where offset distance  $D$  is 9, and orientation  $\theta$  is 0°. The minimum reflectance and the maximum reflectance of each ROI were chosen as the lowest and highest levels of GLCM matrix. The range was divided into eight parts, with each portion being regarded as a level. GLCM matrix at each wavelength resulted in a set of four measurements. As depicted in Fig. 5, a GLCM index (**GI**) at continuous wavelengths could be expressed as a curve across 940–1650 nm. Typical curves for **Git** and **Gln** showed a peak around 1100 nm, similar to features of spectra of **MS** and **MG**. In contrast, **Glh** and **Gla** showed a valley at 1100 nm, indicating that images at this wavelength were less orderly than images at other wavelengths. The highest value of **Git** at 1100 nm implied a large local intensity variation in the NIR image at 1100 nm. Other peaks around 1300 and 1400 nm were apparent for **Git** and **Gln** (Fig. 5). Different measurements with different directions and orientations were tried as variables of multivariate data analysis.

### 3.4. Multivariate data analysis

Features from ROI in calibration set served as input to Eq. (9) and regression matrices (**B**) were obtained. Models were cross validated and tested by independent samples in the prediction set. The results of calibration, cross validation and prediction of all

**Table 2**  
Optimal wavelengths selected from PLSR models of spectral and texture features.

Features	Optimal wavelengths (nm)
<b>MS</b>	958, 986, 1044, 1116, 1135, 1207, 1279, 1457
<b>DMS</b>	962, 1025, 1116, 1140, 1193, 1217, 1318, 1375
<b>MG1</b>	953, 986, 1044, 1116, 1135, 1207, 1270
<b>DMG1</b>	1025, 1121, 1150, 1193, 1217, 1375
<b>MG2</b>	958, 986, 1044, 1116, 1140, 1207, 1279
<b>DMG2</b>	962, 1025, 1121, 1150, 1193, 1217, 1318, 1375
<b>Git</b>	1044, 1140, 1174, 1308, 1400

features are listed in Table 1. The optimal parameters of Gabor filters and GLCM were selected based on the performance of PLSR models. When the algorithm-based PLSR models performed the best, the corresponding parameter sets were used as the optimal parameters for the algorithms:  $\sigma = 10$ ,  $f = 0.1$  for GF1,  $\sigma = 10$ ,  $f = 0.1$ ,  $\theta = 0^\circ$  for GF2,  $D = 9$ ,  $\theta = 0^\circ$ , measurement = *contrast* for GLCM. The raw mean spectra and Gabor filtered spectra showed high similarity in PLSR analysis, while **MG2** performed slightly better than the other features. The prediction of IMF content by **GI** was not as strong as mean spectra of raw ROI and Gabor filtered ROI or the first derivative of mean spectra.

Table 1 shows the regression parameters for models used in selection of optimal wavelengths related to IMF content of pork. Wavelengths corresponding to the first few peak regression coefficients were selected as potential variables, as listed in Table 2. The optimal wavelengths for all the mean spectra were 986, 1044, 1116, and 1207 nm, while the wavelengths of all the first derivative of mean spectra were 1025, 1193, and 1217 nm. Wavelengths of 1044 and 1140 nm were used for GLCM as well as most types of mean spectra. Wavelengths around 960 and 1200 nm were both used as optimal wavelengths for all types of mean spectra and first derivative of mean spectra. Those two wavebands are mainly related to the stretching or deformation vibration of C–H and O–H bonds [28]. The third and second overtones of C–H bonds, which are abundant in fatty acids, resulted absorption



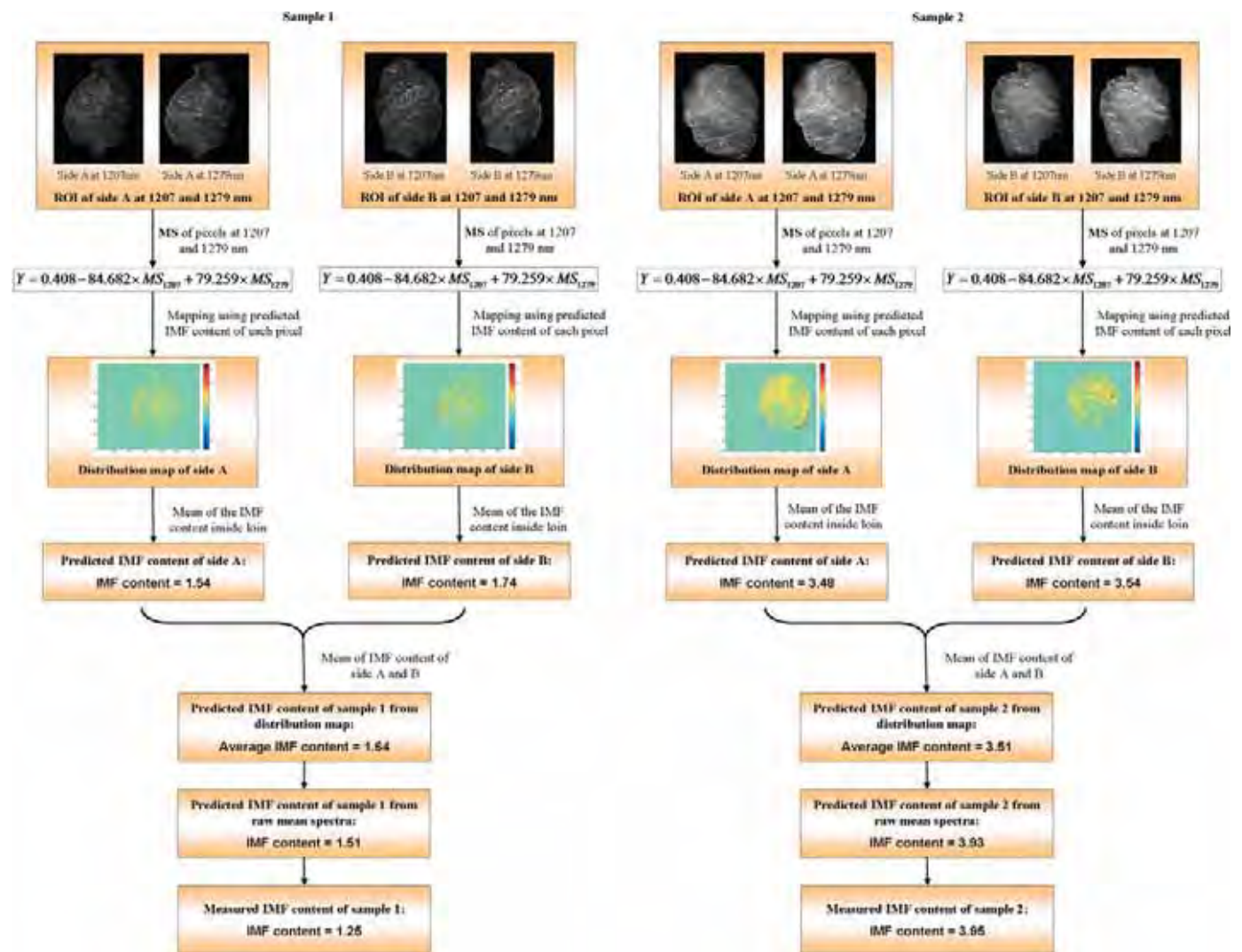
peaks around 930 and 1220 nm [20,28,29]. The frequency around 960 nm, which was derived by stretching vibration of O–H bonds, would be another affect that the band was characterized as an optimal wavelength for all spectral features [20,29,30].

To further lessen the influence of water and lean, a stepwise regression procedure was adopted to simplify the variables at selected wavelengths (Table 2) and further optimize effective wavelengths. The resulting optimal wavelengths, regression coefficients, and results of MLR based on stepwise-selected key

wavelengths are listed in Table 3. Only two key wavelengths were selected for almost all of the features except **DMS**, such that the dimension of hyperspectral data was much reduced in comparison to full waveband-based models. Few wavelengths would help the development of an online determination system for IMF content in intact pork. Wavelengths around 1200 nm were adopted by most features except **GIt**. The rare involvement of spectral feature may cause the outline of 1200 nm for texture feature **GIt**, as 1200 nm is closely related to reflected spectra of fat. A limited number of

**Table 3**  
Results of MLR models of spectral and texture features.

Features	Key wavelengths (nm)	Regression coefficient		Calibration		Cross validation		Prediction	
		$b_0$	$b_i$	$R_c$	RMSEC	$R_{cv}$	RMSECV	$R_p$	RMSEP
<b>MS</b>	1207, 1279	0.408	−84.682, 79.259	0.87	0.52	0.86	0.53	0.85	0.55
<b>DMS</b>	1193, 1217, 1375	−0.063	−405.682, 274.427, −144.927	0.88	0.49	0.88	0.50	0.83	0.58
<b>MG1</b>	1207, 1270	0.538	−3030.206, 2740.172	0.87	0.50	0.86	0.51	0.85	0.53
<b>DMG1</b>	1193, 1217	−0.350	−22904.719, 38033.205	0.89	0.44	0.89	0.44	0.86	0.51
<b>MG2</b>	1207, 1279	0.414	−1210.504, 1125.226	0.87	0.51	0.86	0.51	0.85	0.53
<b>DMG2</b>	1193, 1217	0.376	−6034.862, 5736.390	0.88	0.49	0.88	0.49	0.85	0.52
<b>GIt</b>	1140, 1400	0.787	2.779, −2.998	0.82	0.64	0.81	0.65	0.78	0.76



**Fig. 6.** Visualization of distribution map and the prediction of IMF content of two pork samples.

precise wavelengths would help the development of an effective online determination system for IMF content in intact pork. The performance of key wavelengths-based MLR of **MS**, **DMS**, **MG**, and **DMG** was comparable to the performance of full waveband-based PLSR. The **DMG1** produced the best result, with  $R_c$  of 0.89,  $R_{cv}$  of 0.89, and  $R_p$  of 0.86, using feature **DMG1** at 1193 and 1217 nm. The performance of **GIt** was not as good as the one of other features. However, the MLR model of **GIt** based on key wavelengths outperformed the full waveband-based PLSR model. The further reduction of noise by stepwise analysis may help in data analysis. In this study, a raw mean spectra-based MLR model produced  $R_c$  of 0.87,  $R_{cv}$  of 0.86, and  $R_p$  of 0.85, using spectra at 1207 and 1279 nm. The prediction results of the raw mean spectra-based MLR model showed that IMF content of both calibration and prediction sets was well predicted. While much fewer wavelengths and less data analysis was involved, these results are comparable to the result of  $R_c=0.88$  and  $R_p=0.91$  that was reported by Liu et al. [18]. Considering the complexity of application in practice, mean spectra of raw ROI are suggested as the processing technique for IMF content prediction using hyperspectral imaging.

### 3.5. Visualization of IMF content by distribution map

The MLR model derived from raw mean spectra was used to visualize the distribution of IMF content in pork, by generating both distribution maps of IMF content and the estimated IMF content of pork (Fig. 6). For spectrum of each pixel within a selected ROI, only reflectance at 1207 and 1279 nm was input into MLR model to calculate IMF content of this pixel, i.e. only images at 1207 and 1279 nm were needed for one pork cut. IMF contents of all the pixels inside the ROI yield the distribution map of IMF content. Therefore, the loin portion inside the ROI would affect the profile of the distribution map. Distribution images of two surfaces of one sample were used to generate the predicted IMF content of the sample. Fig. 6 also shows maps of two surfaces of two samples with different measured IMF contents. The ROI of each surface corresponds to one map, showing how IMF content is distributed within the ROI, and in particular, how IMF contents vary drastically between different areas within the same image. In addition, large variation is observed between distribution maps of both sides of a single same sample. The total IMF content of each side (ROI) was generated by averaging the IMF contents of all the pixels within the relative distribution map. The mean of IMF contents of two sides of one sample was used as the predicted IMF content from distribution maps of the sample. The comparison of IMF content of two samples calculated from the raw mean spectrum of ROI, distribution map, and referenced IMF content are depicted in Fig. 6. The error between IMF content from maps and measured values appeared to be slightly larger than the one between IMF content from raw mean spectra and measured values. The prediction model for mapping was built using IMF content measured from the intact pork sample instead of a pixel. Utilization of referenced IMF content from several small portions of the intact pork sample may help to build a more accurate model and therefore enhance the prediction accuracy of the distribution map of IMF content.

## 4. Conclusions

This study investigated the potential of NIR hyperspectral imaging for non-destructive, fast, and objective assessment of IMF content in intact pork. To mine features from numerous data of spectral images, pattern analysis techniques, i.e. the Gabor filter or the GLCM, were applied. Conventional feature extraction methods, including averaging of spectra and first derivative, were

applied as well. Data processing procedure combined by PLSR, stepwise procedure and MLR were used to select effective key wavelengths and establish MLR models. The first derivative of isotropic Gabor filtered mean spectra at 1193 and 1217 nm produced the best result of  $R_c=0.89$ ,  $R_{cv}=0.89$ , and  $R_p=0.86$ . GLCM does not appear to be as effective as Gabor filter and spectral averaging. Considering the feasibility of implementation of such procedures, raw mean spectra are suggested as the most practical feature for prediction of IMF content of intact pork, with little decline in accuracy over the best result by Gabor filter. The wavelengths 1207 nm and 1279 nm were selected as key wavelengths for mean spectra, producing result of  $R_c=0.87$ ,  $R_{cv}=0.86$ , and  $R_p=0.85$ . Distribution map of IMF content using the mean spectra-based MLR model indicated a great potential for detailed IMF inspection in the pork industry.

These promising results demonstrated the great potential of an NIR hyperspectral imaging technique in helping to identify the key wavelengths for IMF content quantification, allowing implementation of a quality assessment component in online inspection systems for intact pork.

## Acknowledgments

The authors gratefully acknowledge the Canadian Centre for Swine Improvement Inc. (CCSI) for providing samples.

## References

- [1] M.S. Brewer, L.G. Zhu, F.K. McKeith, *Meat Sci.* 59 (2001) 153–163.
- [2] E.A. Bryhni, D.V. Byrne, M. Rødbotten, S. Møller, C. Claudi-Magnussen, A. Karlsson, H. Agerhem, M. Johansson, M. Martens, *Meat Sci.* 65 (2003) 737–748.
- [3] J. Brøndum, L. Munck, P. Henckel, A. Karlsson, E. Tornberg, S.B. Engelsen, *Meat Sci.* 55 (2000) 177–185.
- [4] M. Prevolnik, M. Candek-Potokar, D. Skorjanc, S. Velikonja-Bolta, M. Skrlep, T. Znidarsic, D. Babnik, *J. Near Infrared Spec.* 13 (2005) 77–85.
- [5] N. Barlocco, A. Vadell, F. Ballesteros, G. Galieta, D. Cozzolino, *Anim. Sci.* 82 (2006) 111–116.
- [6] B. Savenije, G.H. Geesink, J.G.P. Van der Palen, G. Hemke, *Meat Sci.* 73 (2006) 181–184.
- [7] N. Prieto, R. Roehe, P. Lavín, G. Batten, S. Andrés, *Meat Sci.* 83 (2009) 175–186.
- [8] A.A. Gowen, C.P. O'Donnell, P.J. Cullen, G. Downey, J.M. Frias, *Trends Food Sci. Technol.* 18 (2007) 590–598.
- [9] J. Qiao, M.O. Ngadi, N. Wang, C. Gariépy, S.O. Prasher, *J. Food Eng.* 83 (2007) 10–16.
- [10] L. Liu, M.O. Ngadi, S.O. Prasher, C. Gariépy, *J. Food Eng.* 99 (2010) 284–293.
- [11] D. Barbin, G. ElMasry, D.W. Sun, P. Allen, *Meat Sci.* 90 (2012) 259–268.
- [12] D. Barbin, G. ElMasry, D.W. Sun, P. Allen, *Anal. Chim. Acta* 719 (2012) 30–42.
- [13] Y.Z. Feng, D.W. Sun, *Talanta* 109 (2013) 74–83.
- [14] S. Serranti, D. Cesare, F. Marini, G. Bonifazi, *Talanta* 103 (2013) 276–284.
- [15] G. ElMasry, D.W. Sun, *Meat quality assessment using a hyperspectral imaging system*, in: D.W. Sun (Ed.), *Hyperspectral Imaging for Food Quality Analysis and Control*, Academic Press/Elsevier, USA, 2010, pp. 273–294.
- [16] K. Kobayashi, Y. Matsui, Y. Maebuchi, T. Toyota, S. Nakauchi, *J. Near Infrared Spec.* 18 (2010) 301–315.
- [17] J.P. Wold, M. O'Farrell, M. Hoy, J. Tschudi, *Meat Sci.* 89 (2011) 317–324.
- [18] L. Liu, M.O. Ngadi, S.O. Prasher, C. Gariépy, in: *The Institute of Food Technologists (IFT) Annual Meeting & Food Expo. USA* (2009).
- [19] I. Murray, P.C. Williams, *Chemical principles of near-infrared technology*, in: P. Williams, K. Norris (Eds.), *Near Infrared Technology in the Agricultural and Food Industries*, American Association of Cereal Chemists, Inc., St. Paul, Minnesota, USA, 1987, pp. 17–34.
- [20] J.S. Shenk, M.O. Westerhaus, J.J. Workman. In: D.A. Burns, E.W. Ciurczak (Eds.), *Handbook of Near Infrared Analysis, Practical Spectroscopy Series*, New York, USA, 1992, pp. 383–431.
- [21] Association of Official Agricultural Chemists (AOAC), *Official Methods of Analysis*, 15th ed., AOAC, Washington, DC, 1990.
- [22] L. Liu, M.O. Ngadi, S.O. Prasher, C. Gariépy, *J. Food Eng.* 110 (2012) 497–504.
- [23] D.A. Clausi, M. Jernigan, *Pattern Recogn.* 33 (2000) 1835–1849.
- [24] L. Ma, T. Tan, Y. Wang, D. Zhang, *IEEE Trans. Pattern Recognit. Mach. Intell.*, 25, 1519–1533.
- [25] G. ElMasry, N. Wang, A. El Sayed, M.O. Ngadi, *J. Food Eng.* 81 (2007) 98–107.
- [26] G.K. Naganathan, L.M. Grimes, J. Subbiah, C.R. Calkins, A. Samal, G.E. Meyer, *Sens. Instrum. Food Q. Saf.* 2 (2008) 178–188.
- [27] M.J. Mateo, D.J. O'Callaghan, A.A. Gowen, C.P. O'Donnell, *J. Food Eng.* 99 (2010) 257–262.



- [28] B.G. Osborne, T. Fearn, P.H. Hindle, Applications of near infrared spectroscopy in food and beverage analysis, in: D. Browning (Ed.), *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*, Longman Scientific & Technical, London, 1993, pp. 145–159.
- [29] J.C. Forrest, E.B. Sheiss, M. Morgan, D.E. Gerrard. Pork quality measurement tools-now and in the future, in: *NPPC Pork Quality Summit Proceedings*. Des Moines, IA. 1997, pp. 79–96.
- [30] L. Bokobza., Origin of near-infrared absorption bands, in: H.W. Siesler, Y. Ozaki, S. Kawata, H.M. Heise (Eds.), *Near-Infrared Spectroscopy: Principles, Instruments, Applications*, WILEY-VCH Verlag GmbH, Weinheim, Germany, 2008, pp. 11–41.



## Binding of curcumin to $\beta$ -lactoglobulin and its effect on antioxidant characteristics of curcumin



Ming Li<sup>a,b</sup>, Ying Ma<sup>a,\*</sup>, Michael O. Ngadi<sup>c</sup>

<sup>a</sup> School of Food Science and Engineering, Harbin Institute of Technology, 73 Huanghe Road, Harbin, Heilongjiang 150090, PR China

<sup>b</sup> School of Pharmacy and Food Science, Tonghua Normal University, 950 Yucai Road, Tonghua, Jilin 134001, PR China

<sup>c</sup> Department of Bioresource Engineering, Macdonald Campus, McGill University, Ste-Anne-de-Bellevue, Québec, Canada H9X 3V9

### ARTICLE INFO

#### Article history:

Received 17 October 2012

Received in revised form 14 January 2013

Accepted 27 February 2013

Available online 7 March 2013

#### Keywords:

Curcumin

$\beta$ -Lactoglobulin

Binding

Antioxidant activity

### ABSTRACT

The binding of curcumin (CCM) to bovine  $\beta$ -lactoglobulin ( $\beta$ -Lg) was investigated by Fourier transform infrared and fluorescence. The effect of binding on antioxidant activity of CCM was determined by using ABTS and hydroxyl radical scavenging capacity and total reducing ability. Our results showed that when CCM binds to  $\beta$ -Lg, it lead to a partial change in protein structure. In fact, CCM was bound respectively to two different sites of protein at pH 6.0 and 7.0 via hydrophobic interaction. CCM- $\beta$ -Lg complex was formed by one molecule of protein combining with one molecule of CCM. Moreover, the average distance from one binding site to Trp residues in protein is similar with another. This result suggested that fluorescence resonance energy transfer cannot be used as unique method to study the characteristics of binding of ligands to proteins. The antioxidant activity of CCM might be improved by binding with  $\beta$ -Lg.

© 2013 Elsevier Ltd. All rights reserved.

### 1. Introduction

$\beta$ -Lactoglobulin ( $\beta$ -Lg), a small globular protein with 162 amino acid residues ( $M_r \sim 18,400$ ), is the major whey protein in bovine milk. It is classified as a member of the lipocalin-protein family because of its high affinity to small hydrophobic ligands.  $\beta$ -Lg has a central core consisting of an eight-stranded  $\beta$ -barrel, flanked by a three-turn  $\alpha$ -helix. A final ninth strand forms the greater part of the dimer interface at neutral pH. The  $\beta$ -barrel is formed by two  $\beta$ -sheets, where strands A to D form one sheet and strands E to H form the other (with some participation from strand A, facilitated by a 90° bend at Ser21). The loop EF that connects strands E and F at the open end of the  $\beta$ -barrel acts as a gate (Thompson, Boland, & Singh, 2009, chap. 6). At low pH (<7), the loop EF is closed over the top of the barrel, and binding is inhibited or impossible. At high pH ( $\geq 7$ ), this loop is open, allowing ligands to enter into the hydrophobic core.  $\beta$ -Lg has the ability to bind a large number of small hydrophobic molecules because of its hydrophobic core (Kontopidis, Holt, & Sawyer, 2004). Some studies indicated that  $\beta$ -Lg has three binding sites: the first one is in the  $\beta$ -barrel also known as the central cavity or calyx, the second is the surface cleft which lies between the  $\alpha$ -helix and the surface of the barrel, and the third one is located at the monomer/monomer interface (Harvey, Bell, & Brancaloni, 2007). This carrier property makes it an

attractive candidate to serve as a transporter for delivering important hydrophobic nutrients to improve their bioavailability.

Curcumin (CCM) is a phenolic compound obtained from the dried rhizome of *Curcuma longa* which is a plant widely cultivated in tropical areas of Asia and Central America. Although CCM now is widely used as pigment in food processing, it has been shown to exhibit many bioactivities such as antioxidant, anti-inflammatory, antibacterial, anticancer and against various malignant diseases, diabetes, allergies, arthritis, Alzheimer's disease, and other chronic illnesses (Aggarwal, Sundaram, Malani, & Ichikawa, 2007; Anand, Kunnumakkara, Newman, & Aggarwal, 2007; Anand, Sundaram, Jhurani, Kunnumakkara, & Aggarwal, 2008). In spite of its bioactivities, it is extremely limited in its application due to its low bioavailability related to its insolubility in water as well as its poor absorption and rapid metabolism (Anand et al., 2007). Attempts to improve the bioavailability of CCM have been reported by making complexes with carriers, such as bovine serum albumin (Bourassa, Kanakis, Tarantilis, Pollissiou, & Tajmir-Riahi, 2010), human serum albumin (Mandeville, Froehlich, & Tajmir-Riahi, 2009; Zsila, Bikadi, & Simonyi, 2003),  $\beta$ -lactoglobulin (Mohammadi, Bordbar, Divsalar, Mohammadi, & Saboury, 2009; Snehari, Karakkat, Singh, & Rao, 2010), immunoglobulin (Liu et al., 2008), chitosan (Shelma & Sharma, 2010), phospholipids (Liu, Lou, Zhao, & Fan, 2006) or by formation of liposomes (Li, Braiteh, & Kurzrock, 2005), micelles (Sahu, Kasaju, & Bora, 2008) and encapsulation or nanoparticles (Bisht et al., 2007; Shaikh, Ankola, Beniwal, Singh, & Kumar, 2009; Wang et al., 2008).

\* Corresponding author. Tel.: +86 0451 86282903; fax: +86 0451 86282906.

E-mail address: [maying@hit.edu.cn](mailto:maying@hit.edu.cn) (Y. Ma).

In our experiments,  $\beta$ -Lg was used as a carrier to deliver CCM in order to improve CCM's bioavailability. Except for its high affinity to small hydrophobic ligands, the reason we chose  $\beta$ -Lg is that  $\beta$ -Lg is resistant to pepsin digestion (Sarkar, Goh, Singh, & Singh, 2009). So it can deliver the CCM through the stomach without CCM release due to protein hydrolysis, and then carry the CCM into the intestinal tract. This will be beneficial for CCM to be absorbed intestinally or to treat inflammatory bowel disease more effectively.

Although a few papers about interaction of CCM with  $\beta$ -Lg have been published, some studies are still needed to further investigate on the characteristic of their interaction. Besides, more research is needed in order to understand the effect of binding on the antioxidant activity of CCM. In our experiment we combined the Fourier transform infrared (FTIR) spectra with fluorescence spectra to determine the characteristics of binding of CCM to  $\beta$ -Lg for the first time. After that, we determined the effect of binding on the antioxidant activity of CCM by using ABTS radical scavenging activity, hydroxyl radical scavenging activity and the total reduction capability.

## 2. Materials and methods

### 2.1. Materials

Bovine  $\beta$ -lactoglobulin ( $\beta$ -Lg) (purity > 90%) and 2-deoxyribose was obtained from Sigma Chemical Company. Curcumin (CCM) (analytical grade) was purchased from the Sinopharm Chemical Reagent Co. Ltd., China. ABTS assay kit was obtained from Beyotime Institute of Biotechnology, China. All other materials and reagents were of analytical grade and double distilled water was used for preparing the solutions.

### 2.2. Preparation of $\beta$ -Lg and CCM solution

To obtain protein solution,  $\beta$ -Lg was weighed and dissolved in 0.05 M phosphate buffer solution (pH 6.0 and 7.0). Protein concentration was determined spectrophotometrically from the ultraviolet-visible (UV/Vis) spectra of  $\beta$ -Lg by using a molar absorption coefficient  $\varepsilon_{278\text{ nm}} = 17,600\text{ M}^{-1}\text{ cm}^{-1}$  (Narayan & Berliner, 1997) on a UV/Vis spectrophotometer (754 PC, Shanghai Spectra Instruments Co, Ltd., China). Molecular weight of a  $\beta$ -Lg monomer was taken to be 18,400 Da.

Freshly prepared CCM solution was made up by dissolving the CCM in ethanol to give 2 mM concentration. The exact concentration of CCM was measured by using a molar absorption coefficient  $\varepsilon_{429\text{ nm}} = 55,000\text{ M}^{-1}\text{ cm}^{-1}$  (Mohammadi et al., 2009). The CCM solution was protected from light throughout the experiments.

Complex samples were prepared by mixing  $\beta$ -Lg and CCM solution in varying proportions. The resulting ethanol concentration never exceeded 5% (v/v), which had no appreciable effect on protein structure (Liang, Tajmir-Riahi, & Subirade, 2007).

### 2.3. FTIR spectroscopic measurement

Infrared spectra were carried out on a FTIR spectrometer (spectrum one B, PerkinElmer Inc., USA). All spectra were taken via the KBr pressed disc method. The sample of CCM- $\beta$ -Lg complex was prepared at molar ratio of 1:1 in buffer, and then lyophilized immediately. The difference spectroscopy, second-derivative and curve-fitting were used to analyze the result of FTIR measurement. OriginPro 8.5 software was used to deal with the data.

### 2.4. Fluorescence measurements

Fluorescence measurements were made on a Fluorescence spectrometer (Jasco FP-6500, UK) at room temperature. The fluorescence of CCM binding with  $\beta$ -Lg was measured by keeping CCM concentration at 2  $\mu\text{M}$  and by varying the  $\beta$ -Lg concentration from 0 to 30  $\mu\text{M}$ . The emission spectra were recorded from 430 to 650 nm with an excitation wavelength of 425 nm.

Intrinsic fluorescence of  $\beta$ -Lg was measured at a constant concentration 6  $\mu\text{M}$  in the presence of 0–90  $\mu\text{M}$  CCM. The emission spectra were recorded from 300 to 450 nm with an excitation wavelength of 295 nm.

### 2.5. UV/Vis absorbance measurement

The binding of CCM to  $\beta$ -Lg was also studied by UV/Vis. The path length was 1 cm. CCM- $\beta$ -Lg complex was prepared at molar ratio of 1:1 and the final concentration of CCM was 6  $\mu\text{M}$ . The absorbance spectra of CCM- $\beta$ -Lg complex,  $\beta$ -Lg and CCM were measured respectively from 240 to 310 nm in buffer.

### 2.6. Effect of $\beta$ -Lg on the antioxidant capability of CCM

#### 2.6.1. ABTS radical cation decolorization assay

In our experiments, we used the ABTS assay kit to test CCM and its complexes' antioxidant capacity.  $\text{ABTS}^{\cdot+}$  was produced by reacting ABTS in  $\text{H}_2\text{O}$  with oxidant, stored in the dark at room temperature for 16 h. The  $\text{ABTS}^{\cdot+}$  solution was diluted in 0.05 M sodium phosphate buffer (pH 6.0/7.0). Then, 150  $\mu\text{l}$  of  $\text{ABTS}^{\cdot+}$  solution was added to 50  $\mu\text{l}$  of sample solution. The absorbance was recorded 30 min after mixing. The scavenging capability of test compounds was calculated using the following equation:

$$\text{ABTS}^{\cdot+}\text{scavenging}(\%) = (1 - A_s/A_c) \times 100 \quad (1)$$

where  $A_c$  is absorbance of a control (blank) lacking any radical scavenger and  $A_s$  is absorbance of the remaining  $\text{ABTS}^{\cdot+}$  in the presence of scavenger.

#### 2.6.2. Hydroxyl radical scavenging assay

Ability of samples to scavenge the hydroxyl radical generated by Fenton reaction was measured according to the modified method given by Siriwardhana, Lee, Jeon, Kim, and Haw (2003). The Fenton reaction mixture containing 20  $\mu\text{l}$  of 10 mM  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ , 20  $\mu\text{l}$  of 10 mM EDTA and 20  $\mu\text{l}$  of 10 mM 2-deoxyribose was mixed with 120  $\mu\text{l}$  of 0.05 M sodium phosphate buffer (pH 6.0/7.0) mixed with 20  $\mu\text{l}$  of sample. Thereafter, 20  $\mu\text{l}$  of 10 mM  $\text{H}_2\text{O}_2$  was added before the incubation at 37  $^\circ\text{C}$  for 4 h. 100  $\mu\text{l}$  of 2.8% TCA and 100  $\mu\text{l}$  of 1% TBA were mixed and placed in the boiling water bath for 10 min. After the resultant mixture was cooled to room temperature, the absorbance was recorded at 532 nm.

The percentage of hydroxyl radical scavenged was calculated by using the following equation:

$$\cdot\text{OH}\text{scavenging}(\%) = (1 - A_s/A_c) \times 100$$

where  $A_c$  is the absorbance of the solution without sample,  $A_s$  is the absorbance in the presence of sample.

#### 2.6.3. Total reduction capability

The reducing power of samples was determined by the method of Gulcin (2006). Different concentrations of samples in 100  $\mu\text{l}$  0.05 M sodium phosphate buffer (pH 6.0/7.0) were mixed with potassium ferricyanide [ $\text{K}_3\text{Fe}(\text{CN})_6$ ] (100  $\mu\text{l}$ , 1%). The mixture was incubated at 50  $^\circ\text{C}$  for 20 min. Aliquot (100  $\mu\text{l}$ ) of trichloroacetic acid (10%) was added to the mixture and then was mixed with distilled water (100  $\mu\text{l}$ ) and  $\text{FeCl}_3$  (20  $\mu\text{l}$ , 0.1%), and the absorbance

was measured at 700 nm in a spectrophotometer. Increased absorbance of the reaction mixture indicates an increase of reduction capability.

### 3. Results and discussion

#### 3.1. FTIR spectroscopic measurement

Infrared (IR) spectroscopy is one of the oldest and well established experimental techniques for the analysis of secondary structure of proteins. The vibrations of a structure repeat unit of protein give rise to nine characteristic IR absorption bands, namely, amide A, B, and I–VII. Of these, the amide I and II bands are the most important bands. They are sensitive to the protein secondary structure. The amide I peak position occurs in the region of 1700–1600  $\text{cm}^{-1}$  (mainly C=O stretch), amide II band in the region of 1600–1500  $\text{cm}^{-1}$  (C–N stretch coupled with N–H bending mode). Amide I band is useful and more sensitive to the change of protein secondary structure than amide II (Liu et al., 2008). In this experiment difference spectroscopy and amide I band were carried out to investigate the change of secondary structure of protein in the presence and in the absence of CCM.

As we know that the  $\beta$ -barrel in  $\beta$ -Lg is closed at pH < 7. CCM is not able to get inside the barrel when it interacts with  $\beta$ -Lg. At pH  $\geq 7$  the  $\beta$ -barrel is open and CCM may bind inside, but it may also bind on the surface hydrophobic core. So the pH conditions were just chose at pH 6.0 and 7.0 respectively in our experiments to investigate exactly the binding sites of CCM on protein at acidic and neutral pH respectively. Meanwhile, the existence form of CCM does not change in the pH range 1–7, so the effect of different tautomeric forms of CCM on the spectrum results will not exist in our experiments (Stankovic, 2004).

The spectra of free  $\beta$ -Lg, CCM and CCM- $\beta$ -Lg complex were obtained and normalized in the region of 400–4000  $\text{cm}^{-1}$  (the figures are not shown). The contents of secondary structural compositions of free  $\beta$ -Lg and its CCM complex were calculated according to the curve-fitting of second-derivative procedures using Peak Analyzer in OriginPro 8.5 software and the results are shown in Fig. 1. The free protein was 14%  $\alpha$ -helix (1660–1650  $\text{cm}^{-1}$ ), 34%  $\beta$ -sheet (1642–1610  $\text{cm}^{-1}$ ), 21%  $\beta$ -turn structure (1680–1660  $\text{cm}^{-1}$ ), 17% random coil (1642–1650  $\text{cm}^{-1}$ ), and 14%  $\beta$ -antiparallel (1700–1680  $\text{cm}^{-1}$ ) (Fig. 1A). Upon CCM interaction, only small changes occurred in second structure of  $\beta$ -Lg at pH 6.0 (Fig. 1B). At this pH  $\alpha$ -helix and  $\beta$ -sheet increased from 14% and 34% to 15% and 36%, respectively, with decrease of random and  $\beta$ -antiparallel from 17% and 14% to 15% and 13%, respectively. However, at pH 7.0 the conformational changes of  $\beta$ -Lg interacted with CCM were apparent (Fig. 1C).

At pH 7.0  $\alpha$ -helix increased from 14% to 29%,  $\beta$ -sheet increased from 34% to 38% and small increase in  $\beta$ -turn from 21% to 22%;  $\beta$ -antiparallel decreased from 14% to 11% and random coil was not detected after interaction. The increase in  $\alpha$ -helix structure and  $\beta$ -sheet structures suggested a partial change in protein structure (Liu et al., 2008), which evidenced the formation of CCM- $\beta$ -Lg complex in the binding reaction. Moreover, the difference in the extent of conformational changes between pH 6.0 and 7.0 also gives a proof that CCM binds to different sites at pH 6.0 and 7.0, respectively. Human serum albumin, bovine serum albumin and immunoglobulin A have also been reported to show a partial change in their protein structure when they form complex with CCM (Bourassa et al., 2010; Gupta et al., 2011; Liu et al., 2008; Mandeville et al., 2009).

The results of the spectra also showed that there was hardly any change of peak bands in the range of 1700–1500  $\text{cm}^{-1}$ , which includes the amide I band (1700–1600  $\text{cm}^{-1}$ ; mainly C=O stretch)

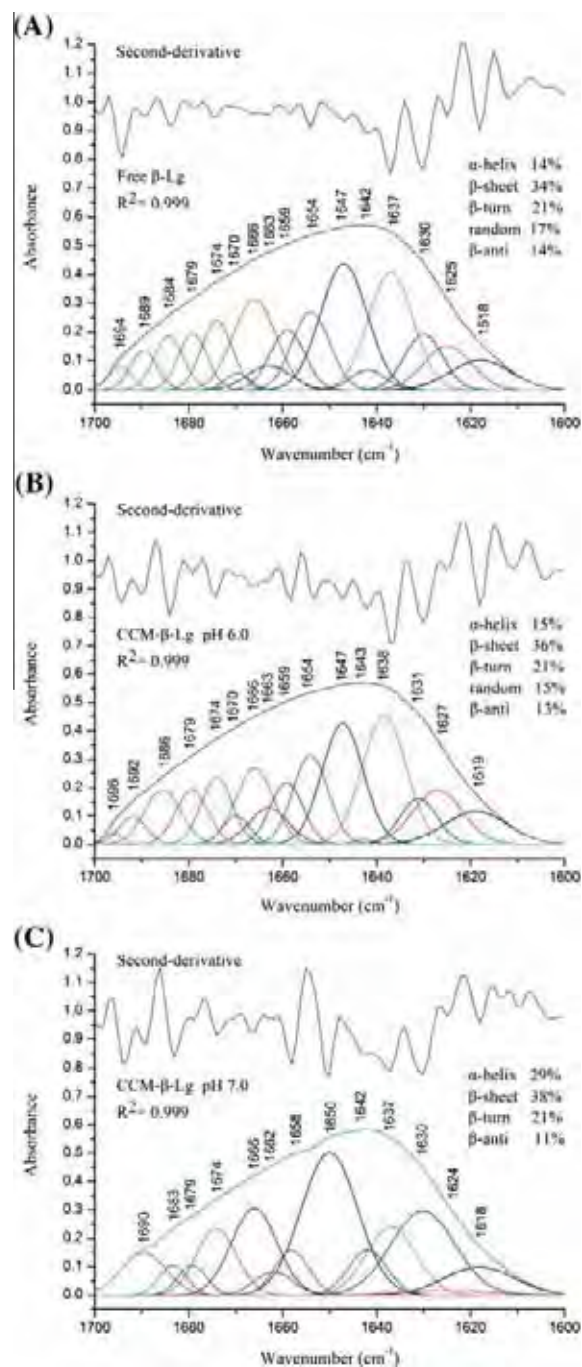


Fig. 1. Curve-fitting with second-derivative of amid I for free  $\beta$ -Lg (A) and its CCM complexes at pH 6.0 (B) and pH 7.0 (C) in the range of 1700–1600  $\text{cm}^{-1}$ .

and amide II band (1600–1500  $\text{cm}^{-1}$ ; C–N stretch coupled with N–H bending mode). The groups of C=O, C–N and N–H are often involved in hydrophilic interactions. However, obvious changes occurred in the range of 3000–2800  $\text{cm}^{-1}$ , which is the spectrum of  $\text{CH}_2$  antisymmetric and symmetric stretching vibrations (Hasni, Bourassa, & Tajmir-Riahi, 2011). The  $\text{CH}_2$  bands of the free  $\beta$ -Lg at 2874, 2933 and 2961  $\text{cm}^{-1}$  shifted to 2875, 2930 and 2968  $\text{cm}^{-1}$  at pH 6.0, to 2876, 2932 and 2970  $\text{cm}^{-1}$  at pH 7.0. Meanwhile, new bands were found at 2923 and 2858  $\text{cm}^{-1}$  (the figures are not shown). The shift of the band is especially obvious at 2961  $\text{cm}^{-1}$ . These results suggest that although CCM contains two hydroxyls, the reaction of CCM with  $\beta$ -Lg is carried out through hydrophobic interactions of polyphenolic rings with



hydrophobic pockets of  $\beta$ -Lg. Similar results have been reported in the reactions of resveratrol, genistein and curcumin with bovine serum albumin (Bourassa et al., 2010).

### 3.2. Fluorescence measurements

Fluorescence measurement is a useful approach to study interaction between ligands and proteins because the fluorophore is sensitive to the polarity of its surrounding environment (Liang et al., 2007). Generally, fluorescence intensity increases as this polarity decrease.

In our study, we found that the fluorescence of CCM showed a low-intensity broad peak at 500 nm in the absence of  $\beta$ -Lg (Fig. 2A and C). When CCM bound to  $\beta$ -Lg, the fluorescence intensity of CCM increased as  $\beta$ -Lg concentration increased. This observation suggested that CCM transferred from the hydrophilic environment of the aqueous solution to a more hydrophobic environment. Furthermore, the increase of fluorescence intensity at pH 7.0 is higher than that at pH 6.0. It indicates that CCM enters into different sites at pH 6.0 and 7.0 and the environment of binding site at pH 7.0 is more hydrophobic.

The data recorded at maximum emission wavelength were used to analyze the binding parameters from the following equation

$$1/\Delta FI = 1/\Delta FI_{\max} + 1/K_a \Delta FI_{\max} [\beta - Lg] \quad (2)$$

where  $\Delta FI$  is the change of the CCM fluorescence in the presence of  $\beta$ -Lg.  $\Delta FI_{\max}$  is the maximal change in fluorescence intensity.  $K_a$  is the binding constant, and  $[\beta\text{-Lg}]$  is the concentration of protein added.

The intensity data were used to plot the double reciprocal plot  $1/[\beta\text{-Lg}]$  versus  $1/\Delta FI$  (Fig. 2B and D). The intercept of the plot on the  $1/\Delta FI$  axis is  $1/\Delta FI_{\max}$ , which was used to calculate the binding constant from the value of the slope in the plot. The binding constant  $K_a$  was estimated to be  $5.23 \times 10^4 \text{ M}^{-1}$  at pH 6.0 and  $8.90 \times 10^4 \text{ M}^{-1}$  at pH 7.0. The linearity of the plot suggests that CCM binds with the  $\beta$ -Lg to form 1:1 complexes (Sahu et al., 2008).

Fluorescence quenching measurement of protein has been widely used to investigate the binding properties of small molecule substances with proteins in solution. At excitation wavelength of 280 nm, both tryptophan (Trp) and tyrosine (Tyr) residues have fluorescence emission, but at excitation wavelength of 295 nm, only the Trp residues show fluorescence emission (Liang et al., 2007).  $\beta$ -Lg contains two Trp residues in positions 19 and 61. Trp<sub>19</sub> is located at the bottom of calyx, a very hydrophobic region, while Trp<sub>61</sub> is located near the aperture of the barrel which renders this residue exposed to the hydrophilic environment (Harvey et al., 2007). Furthermore, Trp<sub>61</sub> is in proximity of Cys<sub>66</sub>–Cys<sub>160</sub> disulfide bridge which is a strong quencher of indole fluorescence. So the fluorescence of  $\beta$ -Lg at excitation wavelength of 295 nm is almost contributed by Trp<sub>19</sub>. Fig. 2E and F shows the change of fluorescence of  $\beta$ -Lg in the presence of CCM with an excitation wavelength of 295 nm. The fluorescence intensity of  $\beta$ -Lg decreased gradually as the increase of CCM concentration. The quenching data were analyzed according to the Stern–Volmer equation:

$$F_0/F = 1 + K_q \times \tau_0 \times [\text{CCM}] = 1 + K_{sv} \times [\text{CCM}] \quad (3)$$

where  $F_0$  and  $F$  are the fluorescence intensities in the absence and presence of CCM.  $[\text{CCM}]$  is CCM concentration.  $K_q$  is the biomolecular quenching rate constant.  $\tau_0$  is the lifetime of fluorophore without quencher ( $\tau_0 = 10^{-8} \text{ s}$ ) (Mohammadi et al., 2009).  $K_{sv}$  is the Stern–Volmer quenching constant.

The Stern–Volmer plot for  $\beta$ -Lg fluorescence quenching by CCM is shown in Fig. 2E and F top insert. The plot of  $F_0/F$  versus  $[\text{CCM}]$  was found to be linear in both cases. The values of  $K_{sv}$  and  $K_q$  are shown in Table 1. The values of  $K_q$  at pH 6.0 and 7.0 are much higher than maximal dynamic quenching constant

( $2.0 \times 10^{10} \text{ M}^{-1} \text{ s}^{-1}$ ) (Liang et al., 2007; Mohammadi et al., 2009). So the main mechanism of quenching is static and CCM– $\beta$ -Lg complex is formed.

For the static quenching, the binding constant  $K_s$  and binding number  $n$  can be calculated according the following equation

$$\log[(F_0 - F)/F] = \log K_s + n \log[\text{CCM}] \quad (4)$$

where  $F_0$ ,  $F$ ,  $[\text{CCM}]$  are the same as the parameters in Stern–Volmer equation. The linear plot of  $\log[(F_0 - F)/F]$  as a function of  $\log[\text{CCM}]$  is given in the lower insert of Fig. 2E and F. From the slope and intercept of the plot, the values of  $n$  and  $K_s$  were obtained, which were shown in Table 1. The value of  $n$  is about 1 at both pH respectively. It suggested that CCM– $\beta$ -Lg complex was formed by one molecule of protein combining with one molecule of CCM. The values of  $K_q$  and  $n$  obtained by Mohammadi et al. (2009) were  $2.49 \times 10^{12} \text{ M}^{-1} \text{ s}^{-1}$  and 0.85. Their conclusions are same with ours.

In addition, some researches about the interactions of other phenolic compound with  $\beta$ -Lg or CCM with other proteins also showed the similar results. The values of  $K_q$  and  $n$  obtained by Liang et al. (2007) were  $3.6 \times 10^{13} \text{ M}^{-1} \text{ s}^{-1}$  and 1.2 in the interaction of resveratrol with  $\beta$ -Lg. It suggested that one molecule of  $\beta$ -Lg combined with one molecule of resveratrol and that the mechanism of quenching is static and resveratrol– $\beta$ -Lg complex is formed. In the studies of interactions of CCM with bovine serum albumin or human serum albumin, the value of  $n$  is also about 1 (Bourassa et al., 2010; Mandeville et al., 2009). However, Liu et al. (2008) obtained that the different concentrations of CCM have an effect on the number of binding sites of CCM on immunoglobulin. When the molar concentrations of CCM increased from 1.664 to 19.61  $\mu\text{M}$  the number of binding sites was 2 and 4.

### 3.3. Fluorescence resonance energy transfer (FRET)

FRET is an electrodynamic phenomenon that occurs between a donor molecule in the excited state and an acceptor molecule in the ground state (Lakowicz, 2006). The donor molecules typically emit at shorter wavelengths that overlap with the absorption spectra of the acceptor. FRET is often used to study the characteristic of interaction between small molecule substances and proteins. By this method the distance between ligands and Trp residues in protein can be obtained, which may contribute to determine the binding site of ligands in protein (Lange, Kothari, Patel, & Patel, 1998). The fluorescence quenching of  $\beta$ -Lg after binding with CCM indicated the occurrence of energy transfer between CCM and  $\beta$ -Lg. According to the Förster's theory, the efficiency of energy transfer,  $E$ , is described by the following equation

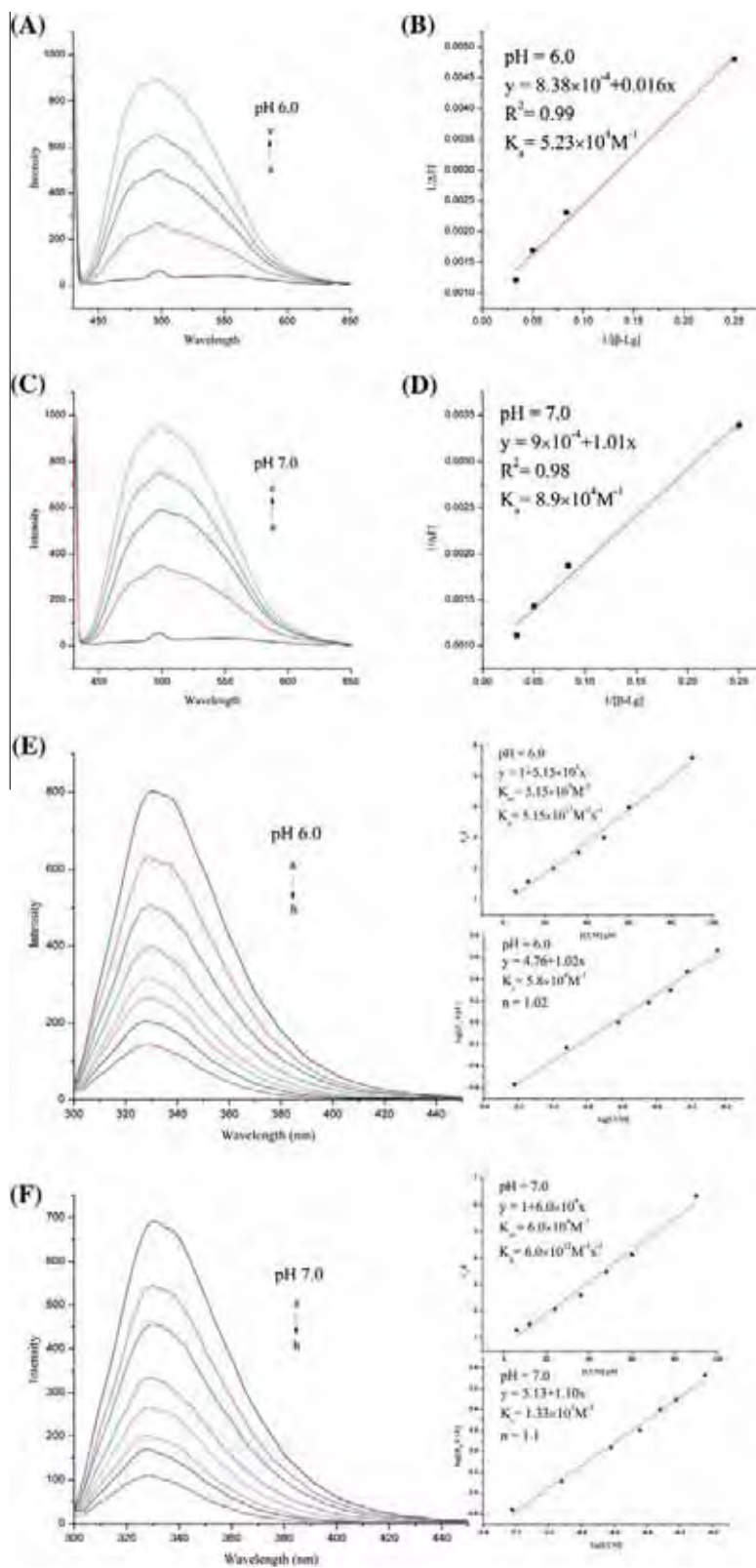
$$E - F/F_0 = R_0^6/(R_0^6 + r^6) \quad (5)$$

where  $F$  and  $F_0$  are the fluorescence intensity of  $\beta$ -Lg in absence and presence of equal amount of CCM,  $r$  is the distance from the CCM bound on  $\beta$ -Lg to the Trp residue, and  $R_0$  is the Förster critical distance which can be calculated by the following equation

$$R_0^6 = 8.8 \times 10^{23} K^2 N^{-4} \Phi J \quad (6)$$

where  $K^2$  is a factor describing the relative orientation in space of the transition dipoles of the donor and acceptor.  $K^2$  is usually assumed to be equal to 2/3.  $N$  is the refractive index which is typically assumed to be 1.4 for biomolecules in aqueous solution.  $\Phi$  is the quantum yield of the donor in the absence of acceptor and was taken to be 0.12 (Lange et al., 1998).  $J$  is the overlap integral between the normalized fluorescence emission spectra of the donor





**Fig. 2.** Fluorescence emission spectra of 2  $\mu\text{M}$  CCM in buffer solution pH 6.0 (A) and pH 7.0 (C) in the presence of  $\beta\text{-Lg}$  at different concentrations (a) 0, (b) 4, (c) 12, (d) 20, (e) 30. The excitation wavelength was 425 nm. Double reciprocal linear plot of  $1/\Delta FI$  as a function of  $1/[\beta\text{-Lg}]$  according to Eq. (2) at pH 6.0 (B) and pH 7.0 (D). Fluorescence emission spectra of 6  $\mu\text{M}$   $\beta\text{-Lg}$  in buffer solution pH 6.0 (E) and pH 7.0 (F) in the presence of CCM at different concentrations (a) 0, (b) 6, (c) 12, (d) 24, (e) 36, (f) 48, (g) 60, (h) 90  $\mu\text{M}$ . The excitation wavelength was 295 nm. Insert: plot of  $F_0/F$  versus  $[\text{CCM}]$  as per the Stern–Volmer equation (top);  $\log[(F_0-F)/F]$  vs  $\log[\text{CCM}]$  as per Eq. (3) (bottom).

and the acceptor absorption spectra.  $J$  is given by the following equation

$$J = \sum F(\lambda)\varepsilon(\lambda)\lambda^4\Delta\lambda / \sum F(\lambda)\Delta\lambda \quad (7)$$

**Table 1**

The Stern–Volmer constant  $K_{sv}$ , biomolecular quenching rate constant  $K_q$ , number of CCM binding sites  $n$  and the binding constant  $K_s$  obtained from the fluorescence quenching measured at pH 6.0 and 7.0.

	pH 6.0	pH 7.0
$K_{sv}$ ( $M^{-1}$ )	$5.15 \times 10^4$	$6.0 \times 10^4$
$K_q$ ( $M^{-1} s^{-1}$ )	$5.15 \times 10^{12}$	$6.0 \times 10^{12}$
$n$	1.02	1.1
$K_s$ ( $M^{-1}$ )	$5.8 \times 10^4$	$1.33 \times 10^5$

In this equation,  $F(\lambda)$  is the fluorescence intensity of the fluorescence donor at wavelength  $\lambda$  and is dimensionless.  $\varepsilon(\lambda)$  is the molar absorption coefficient of the acceptor at wavelength  $\lambda$ .

In our study, The UV/Vis absorption spectra of the CCM in buffer and fluorescence spectra of  $\beta$ -Lg containing same concentration as CCM were recorded. The fluorescence emission of  $\beta$ -Lg excited at 295 nm and ranged from 300 to 450 nm was chosen to obtain the overlapping integral. The overlapping between the fluorescence emission of free  $\beta$ -Lg and absorption spectra of CCM is considerable (the figure is not shown). The efficiency of energy transfer and overlapping integration can be easily calculated using Eqs. (4) and (6). Förster critical distance ( $R_0$ ) can be calculated by Eq. (5). By using the obtained value for  $R_0$  from Eq. (5) and  $E$  from Eq. (4), the  $r$  value can be calculated. As described before,  $\beta$ -Lg has two tryptophan residues (Trp<sub>19</sub> and Trp<sub>61</sub>), so the calculated  $r$  value is the average distance between CCM and these both tryptophan residues. All of the determined parameters for interaction of  $\beta$ -Lg with CCM are reported in Table 2.

$\beta$ -Lg has three binding sites, which are located at the calyx, the surface cleft and the monomer/monomer interface. At low pH (<7), in the structure of  $\beta$ -Lg the loop EF is closed over the top of the barrel, and binding inside the barrel is inhibited or impossible. At high pH ( $\geq 7$ ), this loop is open, allowing ligands to enter into the hydrophobic core. The interaction of CCM with  $\beta$ -Lg is driven by hydrophobic interaction that was proved by FTIR. In our experiment, the value of  $n$  is about 1 at both pH respectively and the values of  $K_a$  and  $K_s$  we obtained were higher at pH 7.0 ( $K_a = 8.90 \times 10^4 M^{-1}$ ,  $K_s = 1.33 \times 10^5 M^{-1}$ ) than that at pH 6.0 ( $K_a = 5.23 \times 10^4 M^{-1}$ ,  $K_s = 5.8 \times 10^4 M^{-1}$ ).

It suggested that one molecule of protein combined with one molecule of CCM and that CCM bound on the surface cleft at pH 6.0, but bound into the calyx at pH 7.0. These results also indicated that the calyx is more hydrophobic environment than surface pocket for the higher values of  $K_a$  and  $K_s$ . Lange et al. (1998) reported that retinol binds to the surface pocket rather than the interior cavity in  $\beta$ -Lg since the values of  $r$  are much different at both binding sites. Nevertheless, from our results, the values of  $r$  obtained from FRET were so similar at both pH conditions. This indicated that the FRET method cannot be used as unique technology to determine the ligands binding sites in this case. At pH 6.0,  $R_0$  is 26.01 Å,  $r$  is 32.4 Å and  $E$  is 0.21. At pH 7.0,  $R_0$  is 26.22 Å,  $r$  is 32.56 Å and  $E$  is 0.21. These results are similar to the results obtained by Mohammadi et al. (2009) ( $R_0 = 25.9$  Å,  $r = 33.8$  Å, pH = 6.4) and Sneharani et al. (2010) ( $R_0 = 26.8$  Å,  $r = 32$  Å,

**Table 2**

The overlap integral  $J$ , Förster critical distance  $R_0$ , distance from the CCM bound on the  $\beta$ -Lg to the Trp residue  $r$  and the efficiency of energy transfer  $E$  obtained from the FRET at pH 6.0 and 7.0 respectively.

	pH 6.0	pH 7.0
$J/M^{-1}cm^{-1}(nm)^4$	$1.69 \times 10^{14}$	$1.78 \times 10^{14}$
$R_0$ (Å)	26.01	26.22
$r$ (Å)	32.40	32.56
$E$	0.2114	0.2147

pH = 7.0). Although the values of their parameters are similar, the pH condition is different. It indicated that the average distance  $r$  between bound CCM and two tryptophan residues is not much different at both binding sites in the case of the interaction of CCM with  $\beta$ -Lg as well as the efficiency of energy transfer  $E$ .

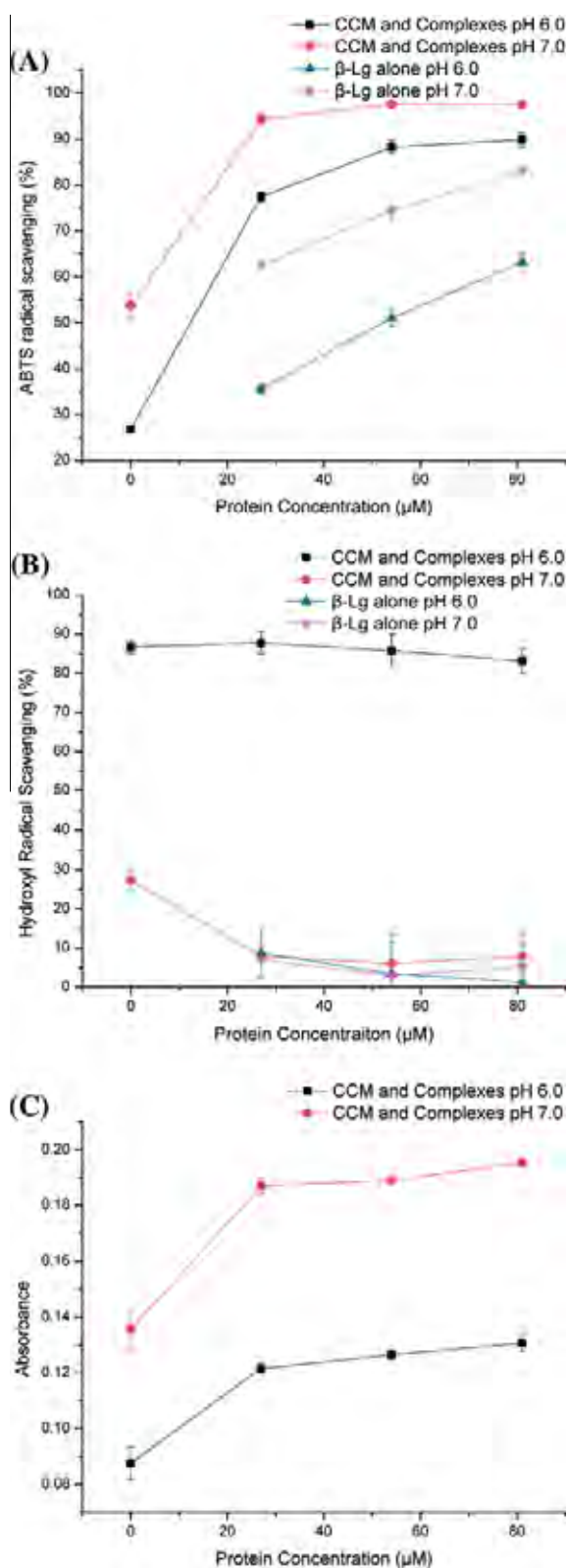
### 3.4. Effect of $\beta$ -Lg on the antioxidant capability of CCM

Recently many studies on the mechanism of CCM's antioxidant activity have been published. There is no doubt that CCM's phenolic hydrogen atoms play an important role in the antioxidant capability. But how does it work has been unclear and under controversy. Litwinienko & Ingold (2004) proposed, on the basis of previous studies and their own experiments, a mechanism named sequential proton loss electron transfer (SPLET) to explain the CCM's antioxidant process. In solvents that support ionization, the initial electron transfer from the ionized keto–enol moiety of CCM to a radical produces a CCM radical fragment which leads one of phenolic hydroxyl groups to lose a proton and then yielding a CCM phenoxy radical. In ionizing solvents, electron-deficient radicals will react with CCM by a rapid SPLET process but in non-ionizing solvents, or in the presence of acid, they will react by a slower hydrogen atom transfer (HAT) process involving one of the phenolic hydroxyl groups.

The effect of binding of CCM to  $\beta$ -Lg on the ABTS radical cation scavenging capability of CCM is depicted in Fig. 3A. In the presence of  $\beta$ -Lg the antioxidant capability of complexes is remarkably higher than CCM alone and has an increasing trend with the increase of protein concentration. Moreover, CCM and its complexes exhibit higher antioxidant capability at pH 7.0 than that at pH 6.0. Although the ABTS radical scavenging capability of CCM is improved when  $\beta$ -Lg is added, it does not prove an effect of  $\beta$ -Lg on CCM's antioxidant activity. This is because  $\beta$ -Lg shows more effective ABTS radical scavenging capability itself than CCM. From the Fig. 3A, we can see that the values of antioxidant capability of  $\beta$ -Lg alone are higher than that of CCM in the same pH condition. Meanwhile, the value increases with the increase of  $\beta$ -Lg concentration, but is lower than that of complexes. The antioxidant capability of  $\beta$ -Lg may be contributed by the sulfhydryl (–SH) that exists on the surface of  $\beta$ -Lg (Liu, Chen, & Mao, 2007). Synergism between the antioxidants in the mixture makes the antioxidant activity not only dependant on the concentration, but also on the structure and the interaction between the antioxidants. So the high activity of complexes may be contributed by both CCM and  $\beta$ -Lg.

Fig. 3B illustrates that there are hardly any effects of  $\beta$ -Lg on the hydroxyl radical scavenging capability of CCM, because there are not apparent changes in the scavenging values of CCM with increase of  $\beta$ -Lg from 0 to 81  $\mu M$  at pH 6.0. Although the scavenging capability of samples is higher at pH 6.0 than that at pH 7.0, it does not prove that this result is contrary to that of ABTS method. This is because the samples were incubated at 37 °C, so this long duration in neutral conditions degraded CCM, which made the samples lose the antioxidant capability (Wang et al., 1997). In this method our results show that  $\beta$ -Lg do not possess an effective scavenging capability.

As shown in Fig. 3C an obvious increase of reduction capability of CCM is observed in the presence of  $\beta$ -Lg. When  $\beta$ -Lg is added there is an increasing trend in total reduction capability of CCM. Moreover, in neutral condition the antioxidant capability of CCM is higher than that in acid condition. This result is consistent with the ABTS cation radical scavenging test. For both methods, the higher antioxidant capability in neutral conditions may be the reason why the SPLET mechanism cannot occur in acidic conditions and the reactions of CCM with radicals involve a slower HAT process. Total reduction capability test is based on the reduction of the  $Fe^{3+}$ /ferricyanide complex to the  $Fe^{2+}$  form. It cannot detect



**Fig. 3.** Effect of different  $\beta$ -Lg concentrations on the antioxidant capability of CCM and CCM- $\beta$ -Lg complexes. (A) ABTS radical scavenging; (B) hydroxyl radical scavenging; (C) total reduction capability. The concentration of CCM is 27  $\mu$ M. The complexes were prepared by CCM interacting with 27, 54 and 81  $\mu$ M  $\beta$ -Lg at the molar ratio of 1:1, 1:2 and 1:3.

compounds that act by HAT, particularly thiols and proteins (Prior, Wu, & Schaich, 2005). So we can now understand the effect of  $\beta$ -Lg on the antioxidant capability of CCM much better. When  $\beta$ -Lg is

added, the reducing power of CCM is improved and increases with the increase of  $\beta$ -Lg concentration. This reason may be that the electron transfer from CCM to  $\text{Fe}^{3+}$  is promoted by the presence of  $\beta$ -Lg and thus improves the rate of SPLET process.

#### 4. Conclusions

The binding of CCM to bovine  $\beta$ -Lg at different pH conditions was investigated by FTIR and fluorescence. The effect of binding on antioxidant activity of CCM was determined by using ABTS radical scavenging, hydroxyl radical scavenging and total reduction capability. Our results showed that when CCM binds to  $\beta$ -Lg, a partial change in protein structure occurred. CCM was bound respectively at two different sites at pH 6.0 and 7.0 via hydrophobic interaction of polyphenolic rings with hydrophobic pockets of  $\beta$ -Lg. At pH 6.0 CCM binds on the surface cleft of  $\beta$ -Lg but it binds into the central cavity at pH 7.0 for the calyx is more hydrophobic environment than surface pocket. CCM- $\beta$ -Lg complex was formed by one molecule of protein combining with one molecule of CCM. Moreover, the average distance from one binding site to Trp residues in protein is similar with another. This result suggested that FRET may not be able to be used as a unique method to study the characteristics of binding of ligands to proteins. The antioxidant activity of CCM might be improved by binding with  $\beta$ -Lg, but it still leads to widely variable results depending on the methods we used in the antioxidant capability test.

#### References

- Aggarwal, B. B., Sundaram, C., Malani, N., & Ichikawa, H. (2007). Curcumin: The Indian solid gold. Molecular targets and therapeutic uses of curcumin in health and disease. *Advances in Experimental Medicine and Biology*, 595, 1–75.
- Anand, P., Kunnumakkara, A. B., Newman, R. A., & Aggarwal, B. B. (2007). Bioavailability of curcumin: Problems and promises. *Molecular Pharmacology*, 4(6), 807–818.
- Anand, P., Sundaram, C., Jhurani, S., Kunnumakkara, A. B., & Aggarwal, B. B. (2008). Curcumin and cancer: An “old-age” disease with an “age-old” solution. *Cancer Letters*, 267(1), 133–164.
- Bisht, S., Feldmann, G., Soni, S., Ravi, R., Karikar, C., & Maitra, A. (2007). Polymeric nanoparticle-encapsulated curcumin (“nanocurcumin”): A novel strategy for human cancer therapy. *Journal of Nanobiotechnology*, 5, 3.
- Bourassa, P., Kanakis, C. D., Tarantilis, P., Pollisiou, M. G., & Tajmir-Riahi, H. A. (2010). Resveratrol, genistein, and curcumin bind bovine serum albumin. *Journal of Physical Chemistry B*, 114(9), 3348–3354.
- Gulcin, I. (2006). Antioxidant and antiradical activities of L-carnitine. *Life Sciences*, 78(8), 803–811.
- Gupta, S. C., Prasad, S., Kim, J. H., Patchva, S., Webb, L. J., Priyadarsini, I. K., et al. (2007). *Natural Products Reports*, 28, 1937–1955.
- Harvey, B. J., Bell, E., & Brancalion, L. (2007). A tryptophan rotamer located in a polar environment probes pH-dependent conformational changes in bovine beta-lactoglobulin A. *The Journal of Physical Chemistry B*, 111(10), 2610–2620.
- Hasni, I., Bourassa, P., & Tajmir-Riahi, H. A. (2011). Binding of cationic lipids to milk beta-lactoglobulin. *Journal of Physical Chemistry B*, 115(20), 6683–6690.
- Kontopidis, G., Holt, C., & Sawyer, L. (2004). Invited review: [beta]-Lactoglobulin: Binding properties, structure, and function. *Journal of Dairy Science*, 87(4), 785–796.
- Lakowicz, J. R. (2006). *Principles of fluorescence spectroscopy*. Springer, pp. 443–472.
- Lange, D. C., Kothari, R., Patel, R. C., & Patel, S. C. (1998). Retinol and retinoic acid bind to a surface cleft in bovine beta-lactoglobulin: A method of binding site determination using fluorescence resonance energy transfer. *Biophysical Chemistry*, 74(1), 45–51.
- Li, L., Braitheh, F. S., & Kurzrock, R. (2005). Liposome-encapsulated curcumin: In vitro and in vivo effects on proliferation, apoptosis, signaling, and angiogenesis. *Cancer*, 104(6), 1322–1331.
- Liang, L., Tajmir-Riahi, H. A., & Subirade, M. (2007). Interaction of  $\beta$ -lactoglobulin with resveratrol and its biological implications. *Biomacromolecules*, 9(1), 50–56.
- Litwinienko, G., & Ingold, K. U. (2004). Abnormal solvent effects on hydrogen atom abstraction. 2. Resolution of the curcumin antioxidant controversy. The role of sequential proton loss electron transfer. *Journal of Organic Chemistry*, 69(18), 5888–5896.
- Liu, H. C., Chen, W. L., & Mao, S. J. T. (2007). Antioxidant nature of bovine milk beta-lactoglobulin. *Journal of Dairy Science*, 90(2), 547–555.
- Liu, A., Lou, H., Zhao, L., & Fan, P. (2006). Validated LC/MS/MS assay for curcumin and tetrahydrocurcumin in rat plasma and application to pharmacokinetic study of phospholipid complex of curcumin. *Journal of Pharmaceutical and Biomedical Analysis*, 40(3), 720–727.

- Liu, Y., Yang, Z., Du, J., Yao, X., Lei, R., Zheng, X., et al. (2008). Interaction of curcumin with intravenous immunoglobulin: A fluorescence quenching and Fourier transformation infrared spectroscopy study. *Immunobiology*, 213(8), 651–661.
- Mandeville, J. S., Froehlich, E., & Tajmir-Riahi, H. A. (2009). Study of curcumin and genistein interactions with human serum albumin. *Journal of Pharmaceutical and Biomedical Analysis*, 49(2), 468–474.
- Mohammadi, F., Bordbar, A.-K., Divsalar, A., Mohammadi, K., & Saboury, A. (2009). Interaction of curcumin and diacetylcurcumin with the lipocalin member  $\beta$ -lactoglobulin. *The Protein Journal*, 28(3), 117–123.
- Narayan, M., & Berliner, L. J. (1997). Fatty acids and retinoids bind independently and simultaneously to beta-lactoglobulin. *Biochemistry*, 36(7), 1906–1911.
- Prior, R. L., Wu, X. L., & Schaich, K. (2005). Standardized methods for the determination of antioxidant capacity and phenolics in foods and dietary supplements. *Journal of Agricultural and Food Chemistry*, 53(10), 4290–4302.
- Sahu, A., Kasoju, N., & Bora, U. (2008). Fluorescence study of the curcumin–casein micelle complexation and its application as a drug nanocarrier to cancer cells. *Biomacromolecules*, 9(10), 2905–2912.
- Sarkar, A., Goh, K. K. T., Singh, R. P., & Singh, H. (2009). Behaviour of an oil-in-water emulsion stabilized by beta-lactoglobulin in an in vitro gastric model. *Food Hydrocolloids*, 23(6), 1563–1569.
- Shaikh, J., Ankola, D. D., Beniwal, V., Singh, D., & Kumar, M. N. (2009). Nanoparticle encapsulation improves oral bioavailability of curcumin by at least 9-fold when compared to curcumin administered with piperine as absorption enhancer. *European Journal of Pharmaceutical Sciences*, 37(3–4), 223–230.
- Shelma, R., & Sharma, C. P. (2010). Acyl modified chitosan derivatives for oral delivery of insulin and curcumin. *Journal of Materials Science-Materials in Medicine*, 21(7), 2133–2140.
- Siriwardhana, N., Lee, K.-W., Jeon, Y.-J., Kim, S.-H., & Haw, J.-W. (2003). Antioxidant activity of *Hizikia fusiformis* on reactive oxygen species scavenging and lipid peroxidation inhibition. *Food Science and Technology International*, 9(5), 339–346.
- Sneharani, A. H., Karakkat, J. V., Singh, S. A., & Rao, A. G. A. (2010). Interaction of curcumin with  $\beta$ -lactoglobulin—stability, spectroscopic analysis, and molecular modeling of the complex. *Journal of Agricultural and Food Chemistry*, 58(20), 11130–11139.
- Stankovic, I. (2004). Curcumin chemical and technical assessment. 61st JECFA © FAO.
- Thompson, A., Boland, M., & Singh, H. (Eds.). (2009). *Milk Proteins – From Expression to Food* (pp. 165–166). Elsevier.
- Wang, X. Y., Jiang, Y., Wang, Y. W., Huang, M. T., Ho, C. T., & Huang, Q. R. (2008). Enhancing anti-inflammation activity of curcumin through O/W nanoemulsions. *Food Chemistry*, 108(2), 419–424.
- Wang, Y.-J., Pan, M.-H., Cheng, A.-L., Lin, L.-I., Ho, Y.-S., Hsieh, C.-Y., et al. (1997). Stability of curcumin in buffer solutions and characterization of its degradation products. *Journal of Pharmaceutical and Biomedical Analysis*, 15(12), 1867–1876.
- Zsila, F., Bikadi, Z., & Simonyi, M. (2003). Molecular basis of the cotton effects induced by the binding of curcumin to human serum albumin. *Tetrahedron: Asymmetry*, 14(16), 2433–2444.



**OVERVIEW PAPER**

## Hyperspectral Imaging for Food Quality and Safety Control

L. Liu and M. O. Ngadi\*

Department of Bioresource Engineering, McGill University, Macdonald Campus. 21,111 Lakeshore Road, Ste-Anne-de-Bellevue, Quebec, Canada. H9X 3V9.

\*Corresponding author: E-mail: [michael.ngadi@mcgill.ca](mailto:michael.ngadi@mcgill.ca)

### ABSTRACT

Hyperspectral imaging (HSI) is an emerging advanced technology in food engineering, especially in food quality and safety analysis and control. The ability of Hyperspectral imaging to obtain both spectral and spatial information of the object of interest makes it a non-destructive, objective, and real-time method for food quality and safety assessment. The present review highlights the fundamentals of HSI; recent advances in the HSI configuration and application of HSI in the food industry; description of image acquisition and processing; and potential for food safety and quality control.

*Keywords:* Hyperspectral imaging; Food quality and safety; Inspection

### INTRODUCTION

With the current need for reduced cost of production, the food industry is facing to a number of challenges including maintenance of high-quality standards and assurance of food safety while avoiding liability issues. Meeting these challenges has become crucial in regards to grading food products for different markets. Food companies and suppliers need efficient, low-cost, quality and safety inspection technologies to enable them satisfy different markets' needs, thereby raising

their competitiveness and expanding their market share.

Quality and safety are usually defined by physical attributes (e.g., texture, color, marbling), chemical attributes (e.g., fat, moisture, protein contents, pH, drip loss), and biological attributes (e.g., total bacterial count). Traditionally, assessment of quality and safety involves human visual inspection and/or chemical and biological determination experiments. Visual inspection is subjective measurement which can be difficult and unreliable and has poor repeatability of results. Chemical and biological determination



experiments are tedious, time-consuming, destructive, and sometimes harmful to the environment. Therefore, it is desirable to develop an accurate, non-destructive, and real-time technique for food quality and safety assessment.

Recently, hyperspectral imaging has been used to evaluate food quality and safety. This technique was originally developed for remote sensing applications as a mean of overcoming the limitations of spectroscopic and machine vision techniques (Goetz *et al.*, 1985). According to Gowen *et al.* (2007), HSI has several merits over RGB, NIR and multispectral imaging. Applications for food quality and safety control include detection of contaminations (Kim *et al.*, 2002, 2004), identification of defects (Xing *et al.*, 2005; Nagata, Tallada & Kobayashi, 2006), analysis of constituents (Qiao *et al.*, 2005), and meat quality evaluation (Qiao *et al.*, 2007). Recently, the technique has become more and more popular due to the consumer demands as well as the challenge of market segmentation and legal restriction. Thus, HSI has shown a strong potential as a powerful detection technique.

### Hyperspectral imaging

Hyperspectral imaging (HSI) combines traditional imaging and spectroscopy technology and can be used to obtain spectral and spatial information of an object of interest over the ultraviolet, visible, and near-infrared spectral regions (200 nm-12  $\mu\text{m}$ ) (Bannon, 2009). Hyperspectral imaging systems provide hyperspectral images consisting of numerous spatial images of the same object at different wavelengths. The hyperspectral image, also called

*hypercube*, is a three-dimensional data cube which is achieved through the superimposition of the spatial images collected by the hyperspectral sensors. These images are composed of vector pixels, and represent the composition and appearance of that particular food sample. Spectra from the data cube of different samples can be compared. Similarity between the image spectra of two samples indicates similarity of chemical composition and physical features. The *hypercube* usually can be constructed in three ways: *area scanning*, *point scanning*, and *line scanning* (Gowen *et al.*, 2007). Due to the presence of conveyor belts (for in-line inspection) in most food processing plants, *line scanning* is the preferred method of image acquisition.

Typical hyperspectral imaging system comprises of hardware and software. The specific configuration may vary depending on the object to be assessed and the technique of image acquisition. Most hardware platform of HSI systems share common basic components as shown in Figure 1: an illumination to provide light source (usually produced by halogen lamps); light irradiation either directly or delivered by optical fiber; detector (e.g. CCD or CMOS camera, InGaAs based array detector) which obtains both spectral and spatial resolution simultaneously; spectrograph to disperse the wavelengths of the light and deliver signals to the photosensitive surface of the detector; an objective lens to adjust the range of light acquisition; an objective table fixed to a conveyor belt to hold and transport the sample; and a computer to compose and store the three-dimensional hypercube.

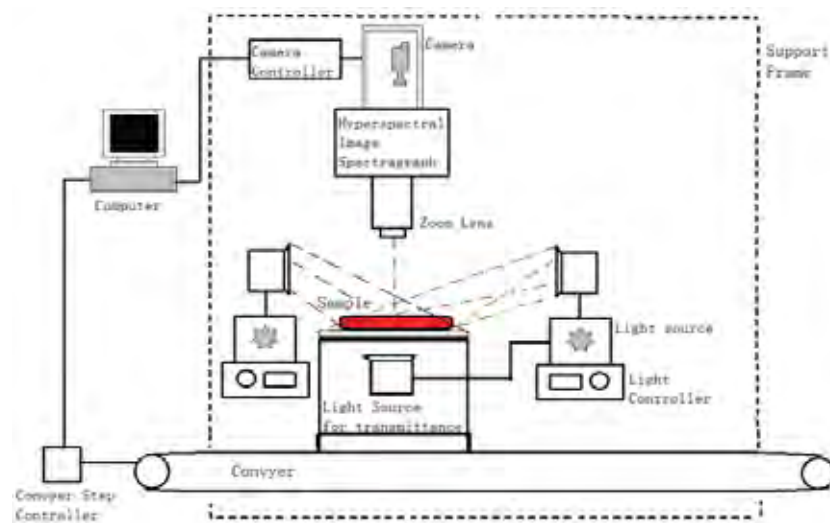


Figure 1. Configuration of a hyperspectral imaging system

Hyperspectral imaging system can be operated either in reflectance or transmittance modes. To acquire images in transmittance mode, thin sample sizes are usually used in order to allow light travelling through the sample. Thick sample can be used in reflectance HSI measurements. Thus, food materials can be inspected as a whole in reflectance mode without the need to make slices. Examples include apples (Peng and Lu, 2008), cucumbers (Ariana and Lu, 2010a), mushrooms (Gowen, Taghizadeh and O'Donnell, 2009), and chickens (Chao *et al.*, 2008).

Hyperspectral imaging system can work in different spectral range such as VIS/NIR range and NIR range. Different detectors are used for different HSI systems due to their sensitivity to the light at different wavelengths. At present, the CCD and CMOS camera (300-1100 nm) are most widely used VIS/NIR detectors in food quality and

safety analysis, while the InGaAs array detector (900-1700 nm, 1000-2200 nm, and 1200-2500 nm) is used for NIR hyperspectral imaging that may provide increased accuracy for assessment and analysis of food quality and safety.

#### Analysis of hyperspectral images

The hypercube contains a mass of information with large dimensionality. The main purpose of hyperspectral data analysis is to reduce the dimensionality and retain the useful data for discrimination or measurement analysis of food quality and safety. Many image processing techniques and chemometric methods can be used to reach the detection goal. Figure 2 illustrates the process of hyperspectral data analysis, including reflectance calibration, segmentation of ROI, image processing and spectral analysis, and classification (qualitative analysis) or prediction (quantitative analysis).

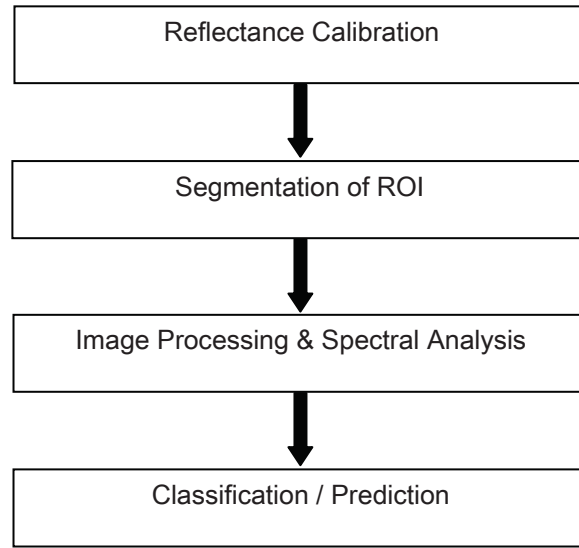


Figure 2. Flow diagram of hyperspectral data analysis process

All hyperspectral images need to be corrected from the dark current of the camera prior to the following data analysis. At the stage of reflectance calibration, the dark response  $D$  and the bright response  $W$  are obtained respectively by covering the lens with the cap and by taking an image from a uniform high reflectance standard (a standard white reference). The corrected reflectance value  $R$  of the original reflected signal  $I$  is calculated on a pixel-by-pixel basis as follows:

$$R = (I - D) / (W - D).$$

After the reflectance calibration, the region of interest (ROI) will be segmented from the background for the further image processing and spectral analysis. Some image processing algorithms can be used to segment ROI, including thresholding, edge detection, filtering, mathematical morphological algorithm, and so on (Martin and Tosunoglu, 2000).

Image processing and spectral analysis will be applied on segmented ROI to extract useful information for

food quality and safety analysis. Image processing algorithms such as GLCM (Qiao *et al.*, 2007), Gabor filters (Liu *et al.*, 2010) and line detection (Liu *et al.*, 2012) can be used to extract image features which can characterize various meat quality attributes at different wavelength bands. Before spectral analysis, the nonchemical biases, i.e. interference signal (baseline drift, particle deviation, surface heterogeneity) should be removed from the spectral information using the spectral processing techniques such as Savitzky-Golay derivative conversion, multiple scattering correction (MSC), and the first or second derivative (Zeaiter, Roger & Bellon-Maurel, 2005). After removing the spectral noise, different statistical procedures such as the analysis of variance (ANOVA), principal component analysis (PCA), linear discriminant analysis (LDA), stepwise regressions, and partial least squares regressions (PLSR) can be used to analyze the spectral data to determine the spectral features which correspond to the various meat quality attributes.

Application of HSI in practice may be limited due to the resulting large and computationally excessive *hypercube*. Thus, it is necessary to extract the characteristic wavelengths by operating qualitative or quantitative analysis for establishment of the relationship between food quality traits and image/spectral characteristics. For qualitative analysis, discriminant analytical tools such as PCA and LDA and machine learning methods such as artificial neural networks (ANNs) and decision trees, are usually employed for classification and evaluation (Camps-Valls & Bruzzone, 2005). Some new methods are also proposed and studied for hyperspectral detection, e.g. the partial least squares discriminant analysis (PLSDA) (Ariana and Lu, 2010a) and spectral information divergence (SID) based classification algorithm (Qin *et al.*, 2009). For quantitative analysis prediction, multivariate analytical tools such as PCA, PLSR, stepwise multi-linear regression (SMLR), are usually employed for chemical content prediction (Cen and He, 2007). PCA and PLSR are the most used and reliable modeling methods. Some advanced techniques such as radial basis function (RBF) (Peng *et al.*, 2008) and PLSDA (Ariana and Lu, 2010a) are also used for prediction and measurement of food quality traits. Table 1 presents a summary of algorithms used for qualitative and quantitative analysis in hyperspectral imaging application for food quality and safety analysis since 2008.

### Application of hyperspectral imaging in food analysis

As an emerging process analytical tool, HSI is well suited for food quality and safety control. Recently, intensive research has been carried with regards to the big potential for hyperspectral imaging application in the food

industry. Table 1 describes the recent advances in the application of hyperspectral imaging on quality and safety analysis of different food products since 2008. An extensive review about the research work reported before 2008 can be found in Gowen *et al.* (2007).

### Fruit

Most of the products studied with hyperspectral imaging are fruits, e.g. apples, citruses, pears, and peaches. The majority of these studies were conducted in reflectance mode and in the VIS-NIR range (400-1100 nm), while some recent research has been carried out in the NIR range (900-1700 nm) (Wang *et al.*, 2009; Sugiyama *et al.*, 2010). For apple, hyperspectral imaging was used to measure quality attributes such as firmness, soluble solids content, and mealiness (Peng and Lu, 2008; Huang and Lu, 2010). These studies demonstrated that hyperspectral scattering technique was potentially useful for nondestructive detection of apple quality attributes. The same spatially resolved diffuse reflectance HSI system was also used to study optical properties of fruits and vegetables including apple, pear, cucumber, tomato (Qin and Lu, 2008). This research reinforced the potential of hyperspectral imaging technique as a convenient attribute classification tools for fruits as well as vegetables.

Qin *et al.* (2009) explored the potential of using hyperspectral imaging to detect citrus canker lesion. The hyperspectral images were processed and classified to differentiate citrus canker lesion from normal and other peel diseased conditions including greasy spot, insect damage, melanose, scab, and wind scar. The analysis method was SID based classification and the overall classification accuracy was 96%. Since this research used full spectral information which was not

desirable for online citrus canker detection, more work needs to be done for optimization of waveband selection.

### Vegetable

A number of papers about mushroom quality detection using HSI have been published since 2008. Hyperspectral imaging was used to detect the bruise damage on white mushrooms (Gowen *et al.*, 2008a) and to identify freeze damage of mushrooms using PCA (Gowen *et al.*, 2009). The quality deterioration of sliced mushrooms during storage was also measured using HSI in terms of moisture content, colour and texture (Gowen *et al.*, 2008b). These studies showed the potential of HSI for damage detection and quality measurement of mushroom. Taghizadeh *et al.* (2010) investigated the shelf life (using parameters including weight loss, color, maturity index) of mushrooms under different packaging polymer films. This research demonstrated that hyperspectral imaging has potential as an analytical tool for evaluation of shelf-life of fresh mushrooms. It also indicated that HSI can be used to evaluate the effect of different packaging solutions especially packaging materials. Hyperspectral imaging was also applied to predict polyphenol oxidase (PPO) activity on mushrooms in the VIS/NIR range (Gaston *et al.*, 2010). The result of this study revealed the possibility of developing a sensor that could rapidly identify mushrooms with a higher likelihood to develop enzymatic browning. This study highlights the utility of HSI in terms of safety and quality management in the food industry.

Ariana and Lu (2010a) developed a VIS-NIR HSI system to classify pickling cucumbers and pickles and detect inner defected pickle pieces. This system combined reflectance mode and transmission mode together and

used a moving transport platform. The results showed the capability of hyperspectral imaging to identify inner defects of cucumber and pickles which was invisible to the naked eyes.

### Meat

Most meat researches related to hyperspectral imaging were performed on pork and beef. Lamb (Kamruzzaman *et al.*, 2010) and ham (EIMasry *et al.*, 2011) have also been investigated.

Hyperspectral imaging was used to classify pork quality groups and predict pork quality traits such as marbling scores (Barbin *et al.*, 2012; Liu *et al.*, 2010, 2012). Barbin *et al.* (2012) studied the grading and classification of three pork quality groups i.e. RFN, PSE, and DFD, using NIR hyperspectral imaging (900-1700 nm). Six wavelengths were selected using the 2nd derivative spectra and PCA was carried out on the selected 6 wavelengths. The high accurate results of this study indicated that pork quality groups could be precisely discriminated using NIR hyperspectral imaging. Instead of three pork quality groups, four pork quality groups (RFN, RSE, PFN, PSE) were classified using VIS/NIR hyperspectral imaging in Liu *et al.* (2010). Automatic and objective measurement methods were established for pork quality classification and quality trait prediction by employing advanced image processing techniques such as Gabor filters and the wide line detector (Liu *et al.*, 2010, 2012). The high accuracy of results indicated the big potential of HSI in meat quality analysis especially when combined with advanced image processing techniques.

The bacterial spoilage process in meat was also studied using a VIS/NIR reflectance HSI system (Peng *et al.*, 2008, 2011). The best prediction result (correlation coefficient=0.95, standard error of prediction (SEP) = 0.30) was obtained using combination of



scattering parameters. These studies demonstrated the great potential of hyperspectral imaging in bacterial activity analysis which is related to food quality change.

### Seafood

The shelf life of salmon was studied by VIS/NIR spectral analysis. Huang *et al.* (2011) applied hyperspectral imaging to predict salmon's storage time. PCA based K-means clustering and MLR were developed to relate hyperspectral data to the storage time and texture change of salmon, respectively. The results indicated that it is possible to predict the texture and storage time using HSI. Barely any research on seafood has been reported in last 2 years. Seafood holds promise as an attractive area for HSI research.

### Biofilm detection

Recently, Jun *et al.* (2010) reported the utilization of macro-scale fluorescence hyperspectral imaging to evaluate the potential detection of pathogenic bacterial biofilm formations on five types of food-contact surface materials: stainless steel, high density polyethylene (HDPE), plastic laminate (Formica), and two variations of polished granite. These materials are commonly used to process and handle food, and sometimes cause biofilm pollution (such as *E. coli* O157:H7 and *Salmonella* biofilm) on food surface. This study showed that the hyperspectral imaging could also be used to develop portable hand-held devices for sanitation inspection of food packaging which has become a big issue for food processing. It was also noted that low cell population density may influence the accuracy of biofilm inspection of food processing surfaces. More studies should be conducted on the HSI biofilm detection in low cell population density.

## CONCLUSION

Hyperspectral imaging (HSI) is developing as a platform technology for food quality and safety analysis in food processing and packaging. Since the information are stored in large data cube i.e. hypercube which may slow down the data processing speed, increasing the efficiency of data analysis methods and identifying the key wavelengths should be the center focus of upcoming studies in order to make HSI more suitable for in plant application.

## REFERENCES

- Ariana, D.P., and Lu, R.F. 2010. Evaluation of internal defect and surface color of whole pickles using hyperspectral imaging. *Journal of Food Engineering*, 96(4): 583-590.
- Bannon, D. (2009). Hyperspectral imaging-cubes and slices. Industry perspective technology focus. *Nature photonics*, 3: 627-629.
- Barbin, D., Elmasry, G., Sun, D.W., and Allen, P. 2012. Near-infrared hyperspectral imaging for grading and classification of pork. *Meat Science*, 90(1): 259-268.
- Camps-Valls, G., and Bruzzone, L. 2005. Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 43: 1351-1362.
- Cen, H., and He, Y. 2007. Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends Food Sci. Technol.*, 18: 72-83.
- Chao, K., Yang, C.C., Kim, M. S., and Chan, D.E. 2008. High throughput spectral imaging system for wholesomeness inspection of chicken. *Applied Engineering in Agriculture*, 24(4): 475-485.
- Cluff, K., Naganathan G.K., Subbiah J., Lu R., Calkins C.R., and Samal A. 2008. Optical scattering in beef steak to predict tenderness using hyperspectral imaging in the VIS-NIR region. *Sensing and Instrumentation for Food Quality and Safety*, 2: 189-196.
- ElMasry, G., Iqbal A., Sun D.W., Allen P., and Ward P. 2011. Quality classification of cooked, sliced turkey hams using NIR hyperspectral imaging system. *Journal of Food Engineering*, 103: 333-344.
- ElMasry, G., Wang, N., and Vigneault, C. 2009. Detecting chilling injury in Red Delicious apple using hyperspectral imaging and neural networks. *Postharvest Biology and Technology*, 52(1): 1-8.
- Gaston, E., Frias, J.M., Cullen, P.J., O'Donnell, C.P., and Gowen, A.A. 2010. Prediction of Polyphenol Oxidase Activity Using Visible Near-Infrared Hyperspectral Imaging on Mushroom (*Agaricus bisporus*) Caps. *Agricultural Food Chemistry*, 58(10): 6226-6233.
- Goetz, A.F.H., Vane, G., Solomon, T.E., and Rock, B.N. 1985. Imaging spectrometry for earth remote sensing. *Science*, 228: 1147-1153.

- Gowen, A.A., Donnell, C.P.O., Cullen, P.J., Downey, G., and Frias, J.M. 2007. Hyperspectral imaging-an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology*, 18(12): 590-598.
- Gowen, A.A., O'Donnell, C.P., Taghizadeh, M., Cullen, P.J., Frias, J.M., and Downey, G. 2008a. Hyperspectral imaging combined with principal component analysis for bruise damage detection on white mushrooms (*Agaricus bisporus*). *Journal of Chemometrics*, 22 (3-4): 259-267.
- Gowen, A., O'Donnell, C., Taghizadeh, M., Gaston, E., O'Gorman, A., Cullen, P. J., Frias, J. M., Esquerre, C., and Downey, G. 2008b. Hyperspectral imaging for the investigation of quality deterioration in sliced mushrooms (*Agaricus bisporus*) during storage. *Sens. Instrum. Food Q. Eval.*, 2(3): 133-143.
- Gowen, A., Taghizadeh, M., and O'Donnell, C.P. 2009. Identification of mushrooms subjected to freeze damage using hyperspectral imaging. *J. Food Eng.*, 93, 7-12.
- Huang, H., Ngadi, M.O., Liu, L., and Benjamin, S. 2011. CIGR Conference.
- Huang, M., and Lu, R.F. 2010. Apple mealiness detection using hyperspectral scattering technique. *Postharvest Biology and Technology*, 58(3): 168-175.
- Jun, W., Kim, M.S., Cho, B.-K., Millner, P.D., Chao, K.L., and Chan, D.E. 2010. Microbial biofilm detection on food contact surfaces by macro-scale fluorescence imaging. *Journal of Food Engineering*, 99 (3): 314-322.
- Kamruzzaman, M., ElMasry G., Sun, D.W., and Allen, P. 2010. Application of NIR hyperspectral imaging for discrimination of lamb muscles. *Journal of Food Engineering*, 104:332-340.
- Kim, M.S., Lefcourt, A.M., Chao, K., Chen, Y.R., Kim, I., and Chan, D.E. 2002. Multispectral detection of fecal contamination on apples based on hyperspectral imagery: Part I. Application of visible and near-infrared reflectance imaging. *Transactions of the ASAE*, 45(6): 2027-2037.
- Kim, I., Kim, M.S., Chen, Y.R., and Kong, S.G. 2004. Detection of skin tumors on chicken carcasses using hyperspectral fluorescence imaging. *Transactions of the ASAE*, 47(5): 1785-1792.
- Liu, L., Ngadi, M.O., Prasher, S.O., and Gariépy, C. 2010. Categorization of pork quality using Gabor filter-based hyperspectral imaging technology. *Journal of Food Engineering*, 99 (3): 284-293.
- Liu, L., Ngadi, M.O., Prasher, S.O., and Gariépy, C. 2012. Objective determination of pork marbling scores using the wide line detector. *Journal of Food Engineering*, 110 (3): 497-504.
- Martin, A., and Tosunoglu, S. 2000. Image processing techniques for machine vision. Florida Conference on Recent Advances in Robotics, Florida Atlantic University, Boca Raton, Florida, May 4-5, 2000.
- Nakariyakul, S., and Casasent, D.P. 2010. Classification of internally damaged almond nuts using hyperspectral imagery. *Journal of Food Engineering*, 99(3): 314-322.
- Naganathan, G.K., Grimes, L.M., Subbiah, J., Calkins, C.R., Samal, A., and Meyer, G.E. 2008. Visible/near-infrared hyperspectral imaging for beef tenderness prediction. *Computers and Electronics in Agriculture*, 64 (2): 225-233.
- Peng, Y., and Lu, R. 2008. Analysis of spatially resolved hyperspectral scattering images for assessing apple fruit firmness and soluble solids content. *Postharvest Biol. Technol.* 48, 52-62.
- Peng, Y. K., and Wang, W. 2008. Prediction of Pork Meat Total Viable Bacteria Count Using Hyperspectral Imaging System and Support Vector Machines. St. Joseph, Michigan: The American Society of Agricultural and Biological Engineers. www.asabe.org. Paper number 7010508, 2008 ASAE Annual Meeting.
- Peng, Y.K., Zhang, J., Wang, W., Li, Y.Y., Wu, J.H., Huang, H., Gao, X.D., and Jiang, W.K. 2011. Potential prediction of the microbial spoilage of beef using spatially resolved hyperspectral scattering profiles. *Journal of Food Engineering*, 102(2): 163-169.
- Qiao, J., Wang, N., Ngadi, M., and Baljinder, S. 2005. Water content and weight estimation for potatoes using hyperspectral imaging. St. Joseph, Michigan: The American Society of Agricultural and Biological Engineers. www.asabe.org. Paper number 053126, 2005 ASAE Annual Meeting.
- Qiao, J., Ngadi, M., Wang, N., Gariépy, C., and Prasher, S. 2007. Pork quality and marbling level assessment using a hyperspectral imaging system. *Journal of Food Engineering*, 83(1): 10-16.
- Qin, J.W., and Lu, R.F. 2008. Measurement of the optical properties of fruits and vegetables using spatially resolved hyperspectral diffuse reflectance imaging technique. *Postharvest Biology and Technology*, 49(3): 355-365.
- Qin, J.W., Burks, T.F., Ritenour, M.A., and Bonn, W.G. 2009. Detection of citrus canker using hyperspectral reflectance imaging with spectral information divergence. *Journal of Food Engineering*, 93 (2): 183-191.
- Sugiyama T., Sugiyama J., Tsuta M., Fujita K., Shibata M., Kokawa M., Araki T., Nabetani H., and Sagara Y. 2010. NIR spectral imaging with discriminant analysis for detecting foreign materials among blueberries. *Journal of Food Engineering*, 101(3): 244-252.
- Taghizadeh, M., Gowen, A., Ward, P., and O'Donnell, C.P. 2010. Use of hyperspectral imaging for evaluation of the shelf-life of fresh white button mushrooms (*Agaricus bisporus*) stored in different packaging films. *Innovative Food Science and Emerging Technologies*, 11 (3): 423-431.
- Wang, W.L., Li, C.Y., Gitaitis, R., Tollner, E.W., and Yoon, S.C. 2009. Detection of Sour Skin Diseases in Vidalia Sweet Onions Using Near-infrared Hyperspectral Imaging. Reno, Nevada: The American Society of Agricultural and Biological Engineers. www. asabe. org. Paper number 096364, 2009 ASAE Annual Meeting.
- Xing, J., Bravo, C., Jancsó, P., Ramon, H., and De Baerdemaeker, J. 2005. Detecting bruises on 'Golden Delicious' apples using hyperspectral imaging with multiple wavebands. *Biosystems Engineering*, 90(1): 27-36.
- Zeaiter, M., Roger, J.M. and Bellon-Maurel, V. 2005. Robustness of models developed by multivariate calibration. Part II: the influence of pre-processing methods. *Trends Anal. Chem.* 24 (5): 437-445.

**Table 1:** Summary of measurement mode, product type, analysis type, wavelength region studied and modelling algorithm employed in papers published on hyperspectral imaging of food since 2008.

Mode	Camera	Product	Wavelength region (nm)	Analysis type	Image processing	Modelling	Author, Year	
Reflectance	CCD	Almond nut	700-1000, 950-1390	Qualitative	Non	Band ratio(BR), Support vector machines (SVM)	Nakariyakul & Casasent, in press	
	CCD	Apple	600-1000	Quantitative	Thresholding(TH)	Partial least squares regression(PLSR), Partial least squares discriminant analysis (PLSDA)	Huang & Lu, 2010	
	CCD	Apple	450-1000	Quantitative	Not-mentioned	Stepwise multi-linear regression(SMLR)	Peng & Lu, 2008	
	CCD	Apple	400-1000	Qualitative	TH	Artificial neural networks (ANN)	EIMasry, Wang & Vigneault, 2009	
	CCD	Apple, Peach	500-1000	Quantitative	TH	Manual analysis	Qin & Lu, 2008	
	CCD	Beef	400-1000	Quantitative	Co-occurrence matrix analysis, PCA	Canonical discriminant	Naganathan <i>et al.</i> , 2008	
	CCD	Beef	400-1100	Quantitative	MLD	Multi-linear regression (MLR)	Peng <i>et al.</i> , 2011	
	CCD	Beef	496-1036	Quantitative	MLD	SMLR	Cluff <i>et al.</i> , 2008	
	CCD	Chicken	389-744	Qualitative	TH	BR	Chao <i>et al.</i> , 2008	
	CCD	Citrus	400-1100	Qualitative	Geometric factor correction(GFC)	Digital elevation model (DEM)	Gómez-Sanchis <i>et al.</i> , 2008a	
	EMCCD	Citrus	450-930	Qualitative	TH	Spectral information divergence (SID) mapping	Qin <i>et al.</i> , 2009	
	CCD	Mandarin	320-1100	Qualitative	GFC	Linear discriminant analysis(LDA), Classification and regression trees (CART)	Gómez-Sanchis <i>et al.</i> , 2008b	
	CCD	Mushroom	400-1000	Quantitative	Not-mentioned	PLSR	Taghizadeh <i>et al.</i> , 2010	
	CCD	Mushroom	400-1000	Quantitative	TH	PCA	Gaston <i>et al.</i> , 2010	
	CCD	Mushroom	400-1000	Qualitative	TH	PCA	Gowen <i>et al.</i> , 2008a	
	CCD	Mushroom	450-850	Quantitative	Not-mentioned	MLR, Principal components regression (PCR)	Gowen <i>et al.</i> , 2008b	
	CCD	Mushroom	450-950	Qualitative	Interactive Selection	PCA	Gowen <i>et al.</i> , 2009	
	CCD	Pork	400-1100	Quantitative	Not-mentioned	Least square support vector machines (LS-SVM)	Peng <i>et al.</i> , 2008	
	CCD	Pork	400-1100	Quantitative	MLD	MLR	Wang <i>et al.</i> , 2010	
	CCD	Pork	400-1000	Quantitative	Gabor-filter, TH	PCA, K-means clustering, LDA	Liu <i>et al.</i> , 2010	
	Transmittance	CCD	Pork	900-1700	Qualitative	TH	PCA	Barbin <i>et al.</i> , In press
		CCD	Pickling cucumbers and whole pickles	400-740	Qualitative	TH	PLSDA, K-nearest neighbor(KNN)	Ariana & Lu, 2010a
CCD		Salmon	400-1100	Qualitative, Quantitative	TH	PCA, K-means clustering, MLR	Huang <i>et al.</i> , 2011	
CCD		Whole pickles	400-675	Qualitative	TH	PCA	Ariana & Lu, 2010b	
InGaAs		Onion	1000-1600	Qualitative	TH	Manual analysis	Wang <i>et al.</i> , 2009	
InGaAs		Maize	960-1662	Qualitative	TH	PLS-DA	Williams <i>et al.</i> , 2009	
HgCdTe		Strawberry	1000-2498	Qualitative	TH	LDA	Sugiyama <i>et al.</i> , 2010	
InGaAs		Pickling cucumbers and whole pickles	1000-1600	Qualitative	TH	PLSDA, KNN	Ariana & Lu, 2010a	
CCD		Whole pickles	740-1000	Qualitative	TH	PLSDA, KNN	Ariana & Lu, 2010a	
CCD		Egg	675-1000	Qualitative	TH	PCA	Ariana & Lu, 2010b	
Fluorescence	CCD	Egg	550-899	Qualitative	Not-mentioned	PCA	Smith, Lawrence & Heitschmidt, 2008	
	EMCCD	Microbial biofilm formation	421-700	Qualitative	TH	PCA	Jun <i>et al.</i> , 2010	





Postharvest Technology of Horticultural Crops - An Overview from  
 Farm to Fork  
 Adel A. Kader

1-8

Innovative Small-scale Postharvest Technologies for reducing losses in  
 Horticultural Crops  
 Lisa Kitinoja

9-15

Novel Food Technologies for Enhancing Food Security and Safety  
 Hosahalli Ramaswamy

17-25

Value-added Fruit Processing and Human Health – Global Perspectives  
 H. P. Vasantha Rupasinghe, Indu Parmar and Lijuan Yu

27-39

Strategies to Improve the Adoption of Postharvest Handling and Agro-Processing  
 Technologies in Africa  
 Leonides Halos-Kim

41-50

Hyperspectral Imaging for Food Quality and Safety Control  
 L. Liu and M. O. Ngadi

51-59

Analytical and sensory monitoring of food products – needs for  
 achieving quality competence  
 Annan N.T.

61-65

Design and Analysis of Experiments for Post Harvest Management  
 and Value Addition Studies with Applications to Essential Oils  
 Tessema Astatie

67-79

Understanding Spices for Processing  
 E. V. Divakara Sasthy

81-89

Integrating gender and HIV/AIDS into food security initiatives:  
 Policy making 'from the ground up'  
 Claudia Mitchell and Cora-Lee Conway

91-101

"Enset is a Good Thing": Gender and Enset in Jimma Zone, Ethiopia  
 Katie MacEntee, Jennifer Thompson, Sirawdink Fikreyesus and Kemenu Jihad

103-109

Feasibility of Including Coffee Waste into Livestock Feeding System  
 of Southwestern Ethiopia  
 Solomon Demke

111-120

Study on Extraction and Storage Stability of Watermelon Pulp Powder as Food  
 Ingredient  
 Yelcnayet Bekele Tola and Hosahalli S. Ramaswamy

121-128

# Effect of Ripening Stages on Basic Deep-Fat Frying Qualities of Plantain Chips

Ogan Mba, Jamshid Rahimi and Michael Ngadi

*Department of Bioresource Engineering, Macdonald Campus of McGill University, Lakeshore Road, Ste-Anne-de-Bellevue, Quebec H9X 3V9 21111, Canada*

Received: December 19, 2012 / Published: May 20, 2013.

**Abstract:** Plantains (*Musa paradisiaca* L.) are a major food staple in West Africa and are cooked in various forms. The objective of this work was to evaluate the frying characteristics of plantains at different stages of ripening. The plantains used in the study were at the unripe stage through four different ripening stages. The samples were peeled and sliced into 2 mm thickness and blanched in hot water at 70 °C for 3 min. The slices were then deep fried in canola oil at 180 °C. The result showed that ripening stage significantly affected moisture loss and fat absorption profiles of the plantain chips. Fully ripened plantain absorbed up to 34% (db) oil during 4 min frying, much higher than unripe plantain. The oil uptake and moisture loss during frying of the plantain chip samples were modelled using 1st order kinetics. The kinetic parameters including rates of moisture loss and oil uptake varied according to the different stages of ripening. Ripening had a significant effect on the colour lightness (L) of the chips. Similarly, the redness (a), the yellowness (b) and textural characteristics were significantly affected by ripening stage.

**Key words:** Plantain ripening, plantain chips, deep-frying, oil uptake, moisture loss, kinetics.

## 1. Introduction

Plantain (*Musa paradisiaca* L.) is native to Southeast Asia, Africa and South America. It is one of the world's most important food crops. At harvest, the pulp and peel of plantain make up about 62% and 38%, respectively. The pulp of matured green unripe fruit is rich in starch, dietary fiber, potassium, magnesium, iron, phosphorus, calcium, vitamin C and carotene. The iron content of plantain is 100% usable for human consumption [1, 2]. Plantain is a climacteric fruit. Its ripening process continues even after harvesting from the parent plant due to evolution of ethylene. Plantain ripening involves irreversible series of physical, physiochemical, biochemical and organoleptic changes in the fruit. During ripening, the peel changes from green to yellow and finally to dark brown. At the same time, pulp color changes from

creamy white to yellow or yellow-orange. The change in the colour of pulp/peel is due to the activity of polyphenol oxidase which is abundant in plantain [3, 4]. Ripe plantain can be eaten fresh on account of its content of vitamin C and other essential minerals. Plantain ripening changes the osmotic balance between pulp and peel. A significant increase in the sugar content of the ripe pulp has been reported [5].

There are many different ways of processing and consuming plantain in West Africa. Deep-frying is a widespread method of plantain chips production. Plantain chips are a popular snack food item in Africa [6]. Plantains are cooked and dried by immersing in hot edible oil (between 130 °C and 190 °C at atmospheric pressure) for a few minutes. This results in simultaneous heat and mass transfer. As heat is transferred from the oil to the food, moisture is evaporated and oil is absorbed. The frying process results in unique flavor, color and texture attributes which are the main drivers of consumer acceptability

---

**Corresponding author:** Michael Ngadi, Ph.D., research fields: food and bioprocess engineering. E-mail: michael.ngadi@mcgill.ca.



of the products [7]. Some models that describe the characteristics of moisture loss and fat uptake, as well as the relationship between moisture loss and fat uptake in deep-fried food products have been developed [8, 9]. The kinetics of the frying process, including the oil diffusion and absorption has been studied for a range of fried products including potatoes and *Musa* species [10-12].

The combined effect of high rate of metabolism, tropical humid conditions, inadequate postharvest handling systems and poor marketing conditions have resulted in postharvest losses of plantains. Plantain becomes very perishable on ripening, resulting in many fruits being discarded. Hence, the need to process ripe plantain into other finished products such as plantain chips. While there are many studies on fried products and oil deterioration, little attention has been paid to the effect of the variability of the raw material on its frying behaviour and the physicochemical properties of the final product [12]. Therefore, the objective of this work was to evaluate the frying characteristics of stored plantains at different stages of ripening. The result will provide useful information to the millions throughout the world of whom the plantain fruit is critical to their nutritional and economic well-being.

## 2. Materials and Methods

### 2.1 Preparation of Samples

Mature Columbian Plantain fingers at unripe green stage and Canola oil, were purchased from a local grocery store in Montreal, Canada. The plantains were stored and allowed to ripen slowly under ambient conditions of 25 °C and 80%-85% RH until they reached the desired different stages of ripeness. The guide was the banana ripening chart [13]. The samples and the stages of ripening used in this work are described in Table 1. On each frying day, the plantain sample was washed and manually peeled. The pulp was sliced to a thickness of  $2 \pm 0.1$  mm using Rival Precision Slicer model C1105. The diagonal slices

**Table 1** Description of plantain ripening stages.

Ripening stage	Storage day	Peel color index number	Peel color
A	1	1	Green
B	3	3	Green-yellow
C	7	6	All yellow
D	8-9	7	Yellow flecked with brown

were blanched at 70 °C for 3 min in a water bath. Paper towel was used to dry the slices and allowed to completely dry at room temperature.

### 2.2 Deep-Frying Process

Four liters of Canola oil was preheated to  $180 \pm 2$  °C in a temperature controlled bench top fryer (model no: D24527DZ, Délonghi Inc. USA). The blanched plantain slices was divided into 5 different portions and fried for 1, 2, 3 and 4 min. The fifth portion was not fried (0 min) which served as control. Ten slices per sampling were put into sample holders that ensured complete immersion in the hot oil. After each frying time, the chips were well drained of adhering surface oil by shaking and blotted with paper towel. The chips were allowed to cool at room temperature. All the experiments were carried out in triplicates. Polyethylene self-sealing storage bags were used to pack the chip samples before the analyses that followed.

### 2.3 Moisture and Fat Analysis

Fried plantain chips were weighed, and then dried at 105 °C to constant weight in a forced air convection oven (Isotemp 700, Fischer Scientific, Pittsburgh, USA). Moisture content was determined gravimetrically [14]. AOAC [15] standard procedure and Soxhlet method were used for fat extraction and analysis. The dried samples were ground using Sumeet electric grinder. Approximately 5.0 g of each ground sample was used for fat extraction. The fat was extracted using petroleum ether in a Soxhlet extractor (SER 148, Velp Scientifica, Usmate, Italy). The fat content was obtained on dry weight bases for each plantain chip sample as the ratio of mass of oil

extracted to mass of dried samples.

#### 2.4 Colour Measurements

The colour parameters ( $L$ ,  $a$  and  $b$ ) of the plantain chips samples were measured by reflectance using a Konica Minolta colorimeter (model no: CR-300, Konica Minolta Sensing Inc. Osaka, Japan). The instrument was standardized each time before reading measurements using a white ceramic plate. Samples were scanned at 5 different locations to determine the  $L$ ,  $a$  and  $b$  values as the average of the five measurements.

#### 2.5 Texture Evaluation

Textural properties were evaluated instrumentally at room temperature by a compression/puncture test using an Instron Universal Testing Machine (Model 4502, Canton MA, USA). Force vs displacement (F/D) curves were generated with the puncture test at different ripening stages and frying times by mounting the sample on a flat rigid support where the distance between the support and a cylindrical punch was 15 mm [16]. The punch diameter and the crosshead speed were 5 mm min<sup>-1</sup> and 25 mm min<sup>-1</sup>, respectively. The textural parameters Hardness (equals Maximum Force/Maximum deformation) and Crispiness (equals slope of the linear section of the curve) [16, 17] were obtained from the F/D curves. Each measurement was repeated five times and their mean values were used as indicators of the parameters.

#### 2.6 Statistical Analysis/Kinetics Model

The SAS system software (Version 9.2, SAS Institute, Inc., 1999, Cary, NC, USA) was used for statistical analysis. Duncan's multiple range test (DMRT) was used to estimate significant differences among means at a 5% probability level. All analyses were conducted in triplicate. Similarly, a first order kinetics was used to describe oil uptake and moisture loss during deep-fat frying [8]. Recognizing that there is a saturation oil content beyond which minimal oil is absorbed into the product during frying, the following

rate equation was used.

$$\frac{dO}{dt} = -k_{ou}(O - O_e) \quad (1)$$

where,  $O$  is oil uptake in the sample,  $O_e$  is the equilibrium oil content,  $K_{ou}$  is the oil uptake rate constant and  $t$  is the frying time. Integrating, the equation can be adjusted and expressed in the form of oil uptake as follows:

$$O_u = O_{us}(1 - e^{-k_{ou}t}) \quad (2)$$

where,  $O_u$  is the fat uptake, (i.e., the difference between the oil content at a given time and the initial oil content),  $O_{us}$  is saturation oil content (i.e., the difference between equilibrium oil content and the initial oil content).

Since frying was done at high temperature (above boiling point of water), equilibrium moisture content was assumed to be zero (i.e., moisture was completely evaporated). Although in reality equilibrium does not exist in deep fat frying, the system behaves as if there is equilibrium. This is due to the physical changes in the product during frying which restrict moisture and oil transfers. Hence the model described by Debnath et al. [18] was used to describe the kinetics of moisture loss ratio:

$$M^* = 1 - e^{-k_{ml}t} \quad (3)$$

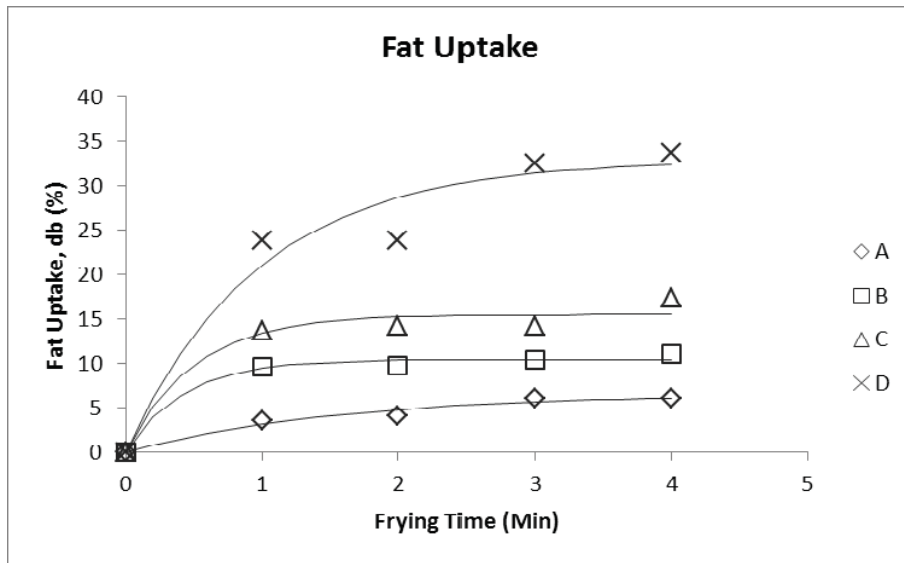
where,  $M^*$  is the moisture ratio (i.e., the ratio of moisture at a given time and initial moisture) and  $K_{ml}$  is the moisture loss rate constant.

The kinetic parameters were obtained using non-linear regression in MATLAB (Version 7.6.0.324 R2008a, The Mathworks, Inc., Natick, MA, USA).

### 3. Results and Discussion

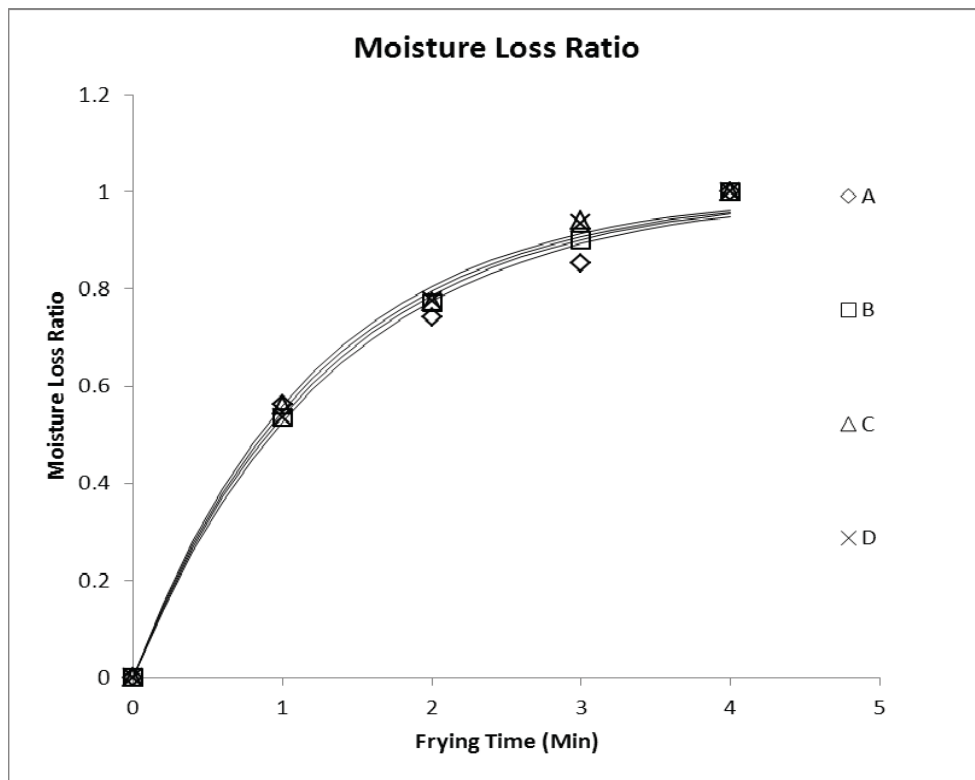
#### 3.1 Fat Uptake/Moisture Loss

The overall influence of plantain ripening on the fat uptake and moisture content of the plantain chips is presented in Figs. 1 and 2. The initial moisture content of mature green plantain was 67.126% ± 0.53% which decreased gradually to a significant 62.274% ± 0.62% by the yellow-brown stage of ripeness. The initial moisture content of the green plantain sample is in



**Fig. 1** Fat uptake trend of unripe and ripe plantain chips during deep-fat frying.

◇ = experimental result plot for unripe green plantain; □ = experimental result plot for plantain at green-yellow stage of ripening; Δ = experimental result plot for plantain at all yellow stage of ripening; × = experimental result plot for plantain at yellow/fleck of brown stage of ripening; Continuous line = Predicted values.



**Fig. 2** Moisture loss trend of unripe and ripe plantain chips during deep-fat frying.

◇ = experimental result plot for unripe green plantain; □ = experimental result plot for plantain at green-yellow stage of ripening; Δ = experimental result plot for plantain at all yellow stage of ripening; × = experimental result plot for plantain at yellow/fleck of brown stage of ripening; Continuous line = Predicted values.

line with the values reported by Agunbiade et al. [19] and Avallone et al. [20]. The decrease in moisture content of the un-fried samples could be due to the reported increase of sugar during ripening from nearly zero at green stage to around 30% in overripe plantain [21]. It was observed that the moisture content of the fried plantain chips continued to decrease as ripening progressed. There was a significant difference in the mean moisture content of each of the fried chips at different ripening stages. As shown in Fig. 1, ripening increases the fat uptake of the plantain chips. There was a significant increase in fat absorption from the initial 0.3% (db) in mature green sample that was not fried reaching up to 34% (db) oil in fully ripened plantain chips after 4 min frying. The increased oil absorption during deep frying of ripe plantains may be attributed to changes in porosity and molecular size redistribution as sugar content of plantains increases. Moreira et al. [22] had shown that porosity and particle size of masa flour used to produce tortilla chips had an effect on the oil uptake during the deep fat frying process.

The average moisture content of the unripe plantain used in this work was 65.0% and was significantly higher than the mean moisture content of 8.01% found in the chips after 4 min of frying. The moisture changes during frying showed the typical extensive decrease with increasing frying time [9, 23]. The trend showed that frying at 1 min led to a rapid moisture decrease (33.78%). This means that the rate of moisture loss of the fried chips increased as ripening progressed and the fruit became softer (Fig. 2). The moisture loss rate was similar as frying time progressed. Generally, the results agree with literature in terms of the moisture content values, oil content values and the effect of temperature on heat and mass transfers. The parameters of the first order kinetic model of fat uptake and moisture loss for different ripening stages of the fried plantain chips is presented in Table 2.

It has been reported that there is a rapid linear

**Table 2 Parameters of first-order kinetic model of fat uptake and moisture loss for different ripening stages of fried plantain chips.**

Ripening stage	$O_{us}$ (% db)	$K_{ou}$ (min <sup>-1</sup> )	$K_{ml}$ (min <sup>-1</sup> )
A	6.611 ± 0.097	0.652 ± 0.493	0.745 ± 0.041
B	10.46 ± 0.388	2.361 ± 0.491	0.773 ± 0.024
C	15.50 ± 0.16	2.015 ± 1.388	0.819 ± 0.028
D	32.93 ± 0.092	1.021 ± 3.357	0.794 ± 0.027

increase in temperature during sensible heating of chips. The crisp was heated from its initial temperature to the boiling point of water. As the oil content plateaus, the moisture content decreased. A threshold moisture value was achieved when the moisture evaporation rate became negligible and the chips were sensibly heated to the frying oil temperature [24, 25]. Diaz et al. [12] reported that oil uptake is determined by the variety of plantain used, irrespective of processing time and temperature. Oil uptake in the case of plantain does not appear to be correlated to the raw material water content as it was previously described for potato chips [26, 27] and cassava chips [28]. Nevertheless, a positive correlation was observed in this work between the amount of oil uptake and the weight of moisture removed. Oil uptake was influenced by stage of ripening and frying time. This result might be explained by an indirect effect associated with the rheological behaviour of ripe plantain, the firmness of which decreased as ripening progressed. Previous studies have also shown that oil uptake was related to the adhesion of oil to the chip surface after removal from the hot oil [22, 29].

### 3.2 Colour/Texture Properties

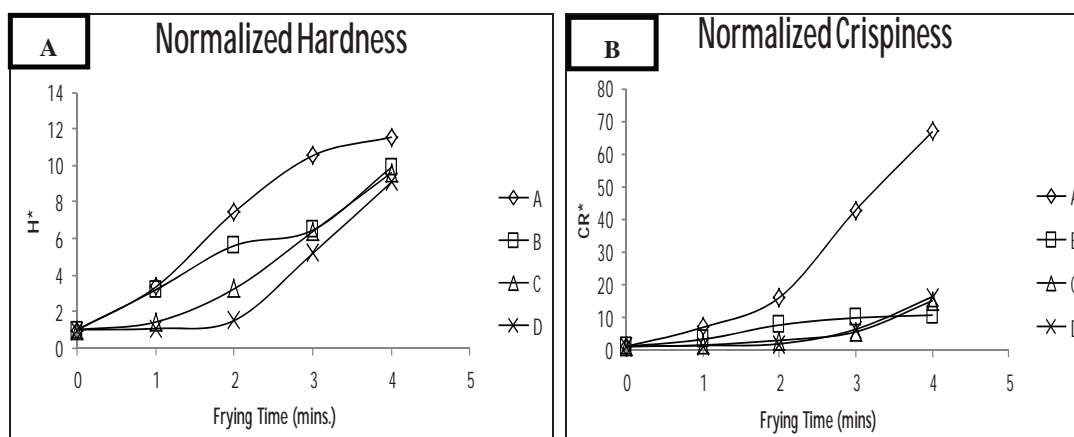
The physical characteristics of the plantain chips, namely colour is shown in Table 3, while texture parameters are shown in Figs. 3a and 3b. In Table 3, the lightness ( $L$ ), redness ( $a$ ), yellowness ( $b$ ), and metric chroma ( $C^*$ ) values are shown. It was observed that at all the ripening stages evaluated the chips had bright colours ( $L > 50$ ). However, the lightness of the fried chips significantly decreased as ripening progressed. The mature “green” samples (A) had a

**Table 3** Colour characteristics of the plantain chips.

Ripening Stages	<i>L</i>	<i>a</i>	<i>b</i>	<i>C</i> *
A	68.01 ± 1.49 <sup>a</sup>	4.57 ± 1.08 <sup>c</sup>	42.13 ± 2.14 <sup>b</sup>	42.74 ± 1.03 <sup>b</sup>
B	65.63 ± 3.46 <sup>b</sup>	6.32 ± 1.73 <sup>b</sup>	42.57 ± 3.08 <sup>b</sup>	43.40 ± 0.90 <sup>b</sup>
C	62.57 ± 4.09 <sup>c</sup>	8.60 ± 1.83 <sup>a</sup>	45.79 ± 3.66 <sup>a</sup>	46.92 ± 0.68 <sup>a</sup>
D	60.99 ± 3.26 <sup>d</sup>	8.78 ± 1.15 <sup>a</sup>	44.78 ± 2.87 <sup>a</sup>	46.09 ± 0.49 <sup>a</sup>

Each value represents the mean ± standard deviation; Means within each column with different letters are significantly different

( $P > 0.05$ ).  $C^* = \text{Metric Chroma} = \sqrt{a^2 + b^2}$ .

**Fig. 3** Effect of ripening on (A) hardness and (B) crispiness of plantain chips.

$H^*$  = Ratio of hardness at any time and hardness at time zero;  $CR^*$  = Ratio of crispiness at any time and crispiness at time zero;  $\diamond$  = experimental result plot for unripe green plantain;  $\square$  = experimental result plot for plantain at green-yellow stage of ripening;  $\triangle$  = experimental result plot for plantain at all yellow stage of ripening;  $\times$  = experimental result plot for plantain at yellow/ fleck of brown stage of ripening.

mean *L* value of 68.01 ± 1.49 while the “yellow/trace of brown” samples (D) had a mean score 60.99 ± 3.26. The redness of the chips samples were significantly increasing as ripening progressed up to the fully ripe “yellow” stage (sample C). There was no significant difference in the redness of chips produced from ripening stages C and D which had redness mean score of 8.60 ± 1.83 and 8.78 ± 1.15, respectively. In general, an increase in redness is negative for colour of fried products [9]. The redness of the plantain chips was within the acceptable range ( $a < 10$ ). The more desired color of fried chips is the yellowness. The data also showed that this parameter was significantly affected by ripening. The result further showed that yellowness of the chips from samples A and B (“green” and onset of ripening) were similar but significantly different from ( $P > 0.05$ ) samples C and D (45.79 ± 3.66 and 44.78 ± 2.87). The mean metric

chroma values confirmed this trend in colour change and indicated a tendency toward browning with increased frying time. The colour changes also indicated more Maillard reaction with frying time which utilized the abundant reducing sugars in ripened plantain [30]. The increase in colour intensity is in agreement with the studies that related the colour of French fries and potato chips [9] and plantains [6] to their reducing sugar content.

The hardness and crispiness of the plantain chips were significantly affected by ripening ( $P > 0.05$ ). Texture of plantain chips is an important sensory parameter that determines acceptability and shelf stability. The data showed a progressive decrease in the mean values of hardness, 16.99 ± 2.53 N mm<sup>-1</sup> (sample A) to 4.90 ± 0.15 N mm<sup>-1</sup> (sample D). As shown in Fig. 3a, hardness increased at all ripening stages as frying time increased due to crust



development [8]. The same trend was observed in the crispiness values which came down from  $20.13 \pm 4.28$  (sample A) to  $2.08 \pm 0.09$  (sample D). Fig. 3b showed that mature green plantain chips had a much higher crispness than the ripe plantains. There was no significant difference in the crispness of chips produced from ripening stages C and D. This could be due to increased oiliness as ripened plantain chip samples had much higher oil uptake. The observed textural changes is in agreement with published data which shows that during fruit ripening, the fruit changes from hard to pulpy soft due to breakdown of cell wall structure and polysaccharides and moisture loss due to respiration [3]. The degradation of pectin from cell wall and middle lamella seems to be responsible for tissue softening of plantain. Kajuna et al. [31] also pointed out a decrease of the modulus of elasticity over the ripening period of plantains as the increase in osmotic pressure led to a reduction in turgor pressure in the pulp tissue. This is attributed to the hydrolysis of starch to sugar.

#### 4. Conclusions

This work evaluated the influence of ripening on the fat uptake and water content of stored plantain. The result showed that ripening had a major effect on fat uptake. The fat content rose from the initial 0.36% to 34% after 4 min of frying fully ripened samples. The moisture loss positively correlated with the fat uptake. The *L*, *a* and *b* color changes of plantain chips were affected by ripening especially from the fully ripe stage. Ripening also significantly affected the textural parameters evaluated. This work established a relationship between plantain ripening and behaviour during frying. This enables utilization of the ripened plantain for chips production rather than allow ripened fruits to waste.

#### References

- [1] P.I. Akubor, E.E. Adejo, Physicochemical, microbiological and sensory changes in stored plantain chips, *J. Plant Foods for Human Nutrition* 55 (2000) 139-146.
- [2] S. Suntharalingam, G. Ravindran, Physical and Biochemical properties of green banana flour, *Plant Foods for Human Nutrition* 43 (1993) 19-27.
- [3] V. Prasanna, T.N. Prabha, R.N. Tharanathan, Fruit ripening phenomena, an overview, *Critical Reviews in Food Science and Nutrition* 47 (2007) 1-19.
- [4] J.O. Offem, O.O. Thomas, Chemical changes in relation to mode and degree of maturation of plantain (*Musa paradisiaca*) and banana (*Musa sapientum*) fruits, *Food Research Inter.* 26 (3) (1993) 187-193.
- [5] J. Marriot, M. Robinson, S.K. Karikari, Starch and sugar transformation during the ripening of plantains and bananas, *J. Sci. Food Agric.* 32 (1981) 1021-1026.
- [6] C.A. Onyejebu, A.O. Olorunda, Effects of raw materials, processing conditions and packaging on the quality of plantain chips, *J. Sci. Food Agric.* 68 (1995) 279-283.
- [7] M. Mellema, Mechanism and reduction of fat uptake in deep-fat fried foods, *Trends in Food Science and Technology* 14 (2003) 364-373.
- [8] P.C. Moyano, F. Pedreschi, Kinetics of oil uptake during frying of potato slices: Effect of pre-treatments, *LWT-Food Science and Technology* 39 (2006) 285-291.
- [9] M.K. Krokida, V. Oreopoulou, Z.B. Maroulis, D. Marinou-Kouris, Effect of pre-drying on quality of french fries, *J. Food Engineering* 49 (2001) 347-354.
- [10] A. Odenigbo, J. Rahimi, M. Ngadi, D. Wees, A. Mustafa, P. Seguin, Quality changes in different cultivars of sweet potato during deep-fat frying, *J. Food Process Technol.* 3 (2012) 156.
- [11] W. Ammawath, Y.B. Che-Man, S. Yusof, R.A. Rahman, Effect of variety and stage of fruit ripeness on the physicochemical and sensory characteristics of deep-fried banana chips, *J. Sci. Food Agric.* 81 (2001) 1166-1171.
- [12] A. Diaz, G. Trystram, O. Vitrac, D. Dufour, A.L. Raoult-Wack, Kinetics of moisture loss and fat absorption during frying of different varieties of plantain, *J. Sci. Food Agric.* 79 (1999) 291-299.
- [13] J. Conie, M. Young, *Banana and Plantain Ripening Manual: A Guide for Small Farmers and Businesses*, Publication of the Banana Board of Jamaica, 2003, pp. 15-18.
- [14] AACC, Moisture Content, in *Approved methods of the American Association of Chemists*, St Paul, MN, 1986.
- [15] AOAC, *Official Methods of Analysis*, Association of Official Analytical Chemists, Washington, DC, 1990.
- [16] F. Pedreschi, P. Moyano, Oil uptake and texture development in fried potato slices, *J. Food Engineering* 70 (2005) 557-563.
- [17] F. Nourian, H.S. Ramaswamy, Kinetics of quality change during cooking and frying of potatoes: Part 1-Texture, *J. Food Process Engineering* 26 (2003) 377-394.

- [18] S. Debnath, K.K. Bhat, N.K. Rastogi, Effect of pre-drying on kinetics of moisture loss and oil uptake during deep fat frying of chickpea flour based snack food, *LWT-Food Science and Technology* 36 (2003) 91-98.
- [19] S.O. Agunbiade, J.O. Olanlokun, O.A. Olaofe, Quality of chips produced from rehydrated dehydrated plantain and banana, *Pakistan Journal of Nutrition* 5 (5) (2006) 471-473.
- [20] S. Avallone, J.A. Rojas-Gonzalez, G. Trystram, P. Bohuon, Thermal Sensitivity of some plantain micronutrients during deep-fat frying, *J. of Food Science* 74 (5) (2009) 339-347.
- [21] D. Mohapatra, S. Mishra, V. Meda, Plantains and their post-harvest uses: An overview, *Stewart Postharvest Review* 5 (4) (2009) 1-11.
- [22] R. Moreira, X. Sun, Y. Chen, Factors effecting oil uptake in tortilla chips in deep-fat frying, *J. Food Engineering* 31 (1997) 485-498.
- [23] L. Kassama, M. Ngadi, Pore development and moisture transfer in chicken meat during deep-fat frying, *J. Drying Technology* 23 (2005) 907-923.
- [24] C.R. Southern, X.D. Chen, M.M. Farid, B. Howard, L. Eyres, Determining internal oil uptake and water content of fried thin potato crisps, *Trans. I. Chem. E.* 78 (2000) 119-125.
- [25] M.M. Farid, X.D. Chen, The analysis of heat and mass transfer during frying of a food using a moving boundary solution procedure, *Heat and Mass Transfer* 34 (1998) 69-77.
- [26] J.H. Makinson, H. Greenfield, M.L. Wong, R.B.H. Wills, Fat uptake during fat frying of coated and uncoated foods, *J. Food Composition Anal.* 1 (1987) 93-101.
- [27] M. Gamble, P. Rice, Effect of initial tuber solids content on final oil content of potato chips, *Lebensm-Wiss u-Technol* 21 (1988) 62-65.
- [28] C.K.V. Nair, C.C. Seow, G.A. Sulebele, Effects of frying Parameters on physical changes of tapioca chips during deep-fat frying, *Int., J. Food Sci. Technol.* 31 (1996) 249-256.
- [29] G. Ufheil, F. Escher, Dynamics of oil uptake during deep-fat frying of potato slices, *Lebensmittel Wissenschaft und Technologie* 29 (1996) 640-644.
- [30] G.I. Onwuka, N.D. Onwuka, The effects of ripening on the functional properties of plantain and plantain based cake, *International Journal of Food Properties* 8 (2) (2005) 1021-1026.
- [31] S. Kajuna, W.K. Bilanski, G.S. Mittal, Textural changes in banana and plantain pulp during ripening, *Journal of Science Food and Agriculture* 75 (1998) 244-250.



---

## ***In vitro* Starch Digestibility and Nutritional Composition of Improved Rice Varieties from Cameroun**

**Amaka M. Odenigbo<sup>1,2</sup>, Sali Atanga Ndindeng<sup>3</sup>, Chijioke A. Nwankpa<sup>1</sup>,  
Noe Woin<sup>3</sup> and Michael Ngadi<sup>1,2</sup>**

<sup>1</sup>*Department of Bioresource Engineering, McGill University, 2111 Lakeshore Road, Ste-Anne-de-Bellevue, Quebec, H9X 3V9, Canada.*

<sup>2</sup>*Department of Human Nutrition and Dietetics, Michael Okpara University of Agriculture, Umudike, Nigeria.*

<sup>3</sup>*IRAD, Yaounde, Cameroon.*

### **Author's contribution**

*This work was carried out in collaboration between all authors. Authors AMO and MN designed the study, performed the statistical analysis, wrote the protocol, and wrote the first and final draft of the manuscript. Authors CAN, SAN and NW contributed to the analyses of the study. All authors read and approved the final manuscript.*

**Received 1<sup>st</sup> June 2013**  
**Accepted 2<sup>nd</sup> October 2013**  
**Published 6<sup>th</sup> November 2013**

**Original Research Article**

---

### **ABSTRACT**

**Aims:** Resistant starch (RS), kinetics of starch digestion, predicted glycemic index (pGI) and nutritional composition were determined in two improved rice varieties from Cameroun.

**Place and Duration of Study:** Department of Bioresource Engineering, McGill University, Canada between December 2012 and March 2013.

**Methodology:** Non-parboiled and parboiled samples of TOX 3145 and NERICA-3 varieties were involved in this study. An *in vitro* enzymatic starch digestion method was applied to measure starch digestibility parameters. Standardized methods were adopted for proximate and mineral contents evaluation.

**Results:** The parboiled samples had significantly higher ( $P < 0.05$ ) resistant starch (8.35 - 11.07%) than the non-parboiled samples (3.81 - 4.84%). The values for pGI among

---

\*Corresponding author: Email: [michael.ngadi@mcgill.ca](mailto:michael.ngadi@mcgill.ca);

samples ranged from 57.57 to 67.78%. Significantly higher values for protein, phosphorus and potassium were found among the parboiled samples ( $P < 0.05$ ). Nutritional composition was positively related to RS while pGI had inverse relationship with protein, ash, fat, phosphorus, potassium and RS.

**Conclusion:** Starch digestibility of these rice varieties was associated to their nutritional composition.

*Keywords:* Starch digestibility; milled rice; parboiling; resistant starch, improved rice variety; nutritional composition.

## 1. INTRODUCTION

Rice has become a staple food and a major source of energy in Cameroun [1]. The increased consumption rate of rice has made the screening of its glycemic index a relevant area of research interest. Glycemic index (GI) is a property of starchy food, which describes the rate of blood glucose absorption after consumption [2].

Milled rice is predominantly, starchy endosperm with a higher glycemic index compared to brown rice that contains the outer bran and germ portions [3]. In the process of milling, removal of only husk from paddy results to production of brown rice while removal of bran and most of the germ layer leads to production of milled rice (white rice).

Nutrients are more concentrated in the bran than the endosperm. It is a known fact that milling reduces the nutritional composition of rice grain, but on the other hand, the milling processes remove a large proportion of the anti-nutrient such as phytate, which might adversely affect the utilization of some of the nutrients [4]. The milled rice could be parboiled (parboiled milled grain) or non-parboiled (raw milled grain). Parboiling process involves three operations namely; soaking, steaming and drying of paddy.

The techniques in parboiling vary from traditional method to more sophisticated procedures [5]. These various parboiling operation differ in degree and intensity of temperature as well as soaking and steaming duration. The grain of parboiled rice is modified by the processes of soaking in hot water, steaming and drying it undergoes and this parboiled grain is not classed as cooked rice.

Thus, parboiled milled rice are grains of paddy which are milled after the process of parboiling whereas, non-parboiled milled rice are grains from milled raw paddy not subjected to parboiling process.

Doesthale et al. [6] studied the effect of milling on nutrient loss in parboiled and non-parboiled rice grain. Their study reported significantly lower milling losses for nutrients in parboiled than non-parboiled rice. Higher concentration of nutrient found in parboiled rice was attributed to nutrient solubilization and migration to the centre of the grain during starch gelatinization which occurs during the parboiling process [7,8].

However, Heinemann et al. [7] pointed the lack of uniformity in commercial parboiling processes in different countries as a hindrance to the conclusion of superior nutritional benefits of parboiled rice.

The benefits of parboiled rice grain over non-parboiled apart from firmer, stronger, less sticky grains, increase of milling recovery and decrease of cooking losses include increase in resistance starch fraction and lower glycemic index [9-12].

Resistant starch (RS) escapes enzymatic hydrolysis in the upper gastrointestinal tract which renders this fraction of starch unavailable to digestion with consequent reduction of post-prandial response. The relationship between RS and glycemic index of food had been described [13,12].

Modern research efforts of both national and international organizations have contributed to the improvement and development of newly improved rice varieties. The New Rice for Africa (NERICA) developed by AfricaRice (formerly West Africa Rice Development Association) are products of hybridization between the cultivated rice species of *Oryza glaberrima* and *Oryza sativa* [14]. In addition to the desirable agronomic characteristics (high yield, early maturity and resistant to drought, pests, diseases) of the improved varieties, the protein content is high [15,16].

The influence of protein content [17,18], moisture content [19], phosphorus contents [20,21,22] and resistant starch [23,24] on starch digestibility have been demonstrated. With the progressive release of newly improved rice varieties, it is relevant to evaluate the nutritional potential and glycemic index of each variety in order to provide optimal nutrition benefits to consumers.

There is paucity of literature data on nutritional composition and starch digestibility of newly improved rice varieties in different processed forms. This study aimed to screen the resistant starch fractions, predicted glycemic index, starch hydrolysis kinetics, proximate and mineral composition of two improved rice varieties grown in Cameroun.

## 2. MATERIALS AND METHODS

### 2.1 Samples

TOX 3145 and NERICA-3 rice varieties are improved lines of *Oryza sativa L* originated from AfricaRice. Samples of freshly harvested paddy of these two varieties were collected from a rice farmer in Ndop, Northwest Region of Cameroun. The varieties were harvested by panicle picking and separated into three portions. A portion from each rice variety was not parboiled (NP), dried to 12% moisture content before milling (NERICA-NP and TOX-NP samples). The other two portions from each variety were parboiled by two different methods, and then dried to moisture content of 14%.

Parboiled TOX 3145 variety was treated by Traditional (TOX-Trad samples) and IRAD-direct (TOX-IRAD) parboiling techniques.

Traditional parboiling was carried out by a farmer's group well known for parboiling rice at a semi-industrial scale in Cameroun. This traditional procedure involved no pre-cleaning of paddy, longer soaking and steaming time (18-20 hrs), soaking temperature of 80 °C, non-uniform distribution of steam during steaming with traditional cooking equipment (barrel drums) and a three stone fire place.



The IRAD-direct parboiling was also carried out by the same farmer's group but the procedure involved an improved direct parboiling technology which has been code named a Uniform-Steam Parboiling system in IRAD (Unpublished report). The duration for soaking was 12 hrs at 80 °C. This technique involved use of an improved steaming equipment (Vessel, stand, steam basket) and improved parboiling stove to ensure uniform distribution of steam.

On the other hand, NERICA-3 variety was parboiled by both subsets of IRAD parboiling process IRAD-direct and IRAD-indirect parboiling techniques.

Direct parboiling technique involves putting the rough rice (paddy) in water, placed on the fire before heated to 80 °C. At 80 °C temperature, the paddy was removed from the fire and left to stand in the water for 12 hr.

For indirect parboiling, water was heated to boiling point (98 °C) and poured on paddy while stirring. The paddy was then left to stand in the water for 12 hr.

Both parboiled and non-parboiled samples were dehusked with Satake rice husker (THU 35A, Satake, Engineering Co. Ltd., Tokyo) and milled by a commercial roll milling facility. The milled rice grains were then brought to McGill University, Canada for analysis.

### **2.1.1 Sample preparation for analysis**

#### *2.1.1.1 Raw samples*

Grains of the rice samples were individually ground using a coffee grinder (SUMEET Multi Grind, India) and passed through a 60 mesh (0.25 mm) sieve (CE Tyler, Ontario, Canada).

#### *2.1.1.2 Cooked samples*

Weight of  $50 \pm 0.05$  mg sample was weighed into capped tube and boiled in tap water (5 mL) for 30 min. Cooked samples were homogenized in cooking water for 1 min using Tissue-Tearor homogenizer (Biospec Products. Inc.) with controlled speed (level 2). The *in vitro* starch digestibility analysis commenced immediately after homogenization [25].

## **2.2 Nutritional Composition**

Nutrient composition was determined by standard methods of the Association of the Official Analytical Chemists [26]. Moisture was estimated by drying in an oven at 105 °C until constant weight; ash contents by combustion of sample at temperature of 550 °C in muffle furnace; protein by nitrogen determination using Leco Nitrogen Analyzer (N x 6.25) while fat determination was by Soxhlet extraction using petroleum ether as solvent. The total carbohydrate content was determined by difference using the following formula:  $100 - (\text{weight in grams} [\text{protein} + \text{fat} + \text{water} + \text{ash}] \text{ in } 100 \text{ g sample})$ .

Concentrations of phosphorus, potassium, iron and magnesium were determined by digestion of ground sample at 340 °C using a digestion mixture of sulfuric acid, peroxide (30%), lithium and selenium. After digestion, phosphorus was estimated colorimetrically at the wavelength of 880 nm on a flow injector analyzer instrument (QuickChem series 8000, Lachat Instruments, CO, USA). Potassium, magnesium and iron were measured with a

flame atomic absorption spectrophotometer (Perkin-Elmer 2380, Norwalk, Connecticut, USA).

### 2.3 Starch Fractions and Predicted Glycemic Index (pGI)

An *in vitro* method based on the procedure of Goni *et al.* [25] was adopted for determination of predicted glycemic index (pGI). The resistant starch (RS), digestible starch (DS) and total starch (TS) were measured according to AOAC 2002.02 using the Megazyme RS kit (Megazyme, Bray, Ireland).

### 2.4 Statistical Analysis

The experimental design consisted of two rice varieties (TOX 3145 and NERICA-3) subjected to three treatments (non-parboiled, two differently parboiled techniques) totalling 6 samples.

Data were expressed as mean±standard deviation of four replicate measurements. Variation in levels of starch fractions among samples were tested by a one way analysis of variance (ANOVA). The difference between means was determined by Scheffe's multiple comparison procedure test ( $P<0.05$ ). Statistical software used was SAS version 9.3 (SAS Institute Inc., Cary, NC, USA). The kinetic parameters were calculated by a nonlinear regression in MATLAB (Version 7.6.0.324 R2008a, The Mathworks, Inc., Natick, MA, USA).

## 3. RESULTS

Proximate composition showed that protein and ash contents among samples ranged between 8.97 -14.98% and 0.52-1.19% dry matter, respectively (Table 1). The fat content among samples was statistically similar ( $P>0.05$ ).

TOX-NP sample had the lowest values for fat (0.39%), protein (8.97%) and highest carbohydrate content (90.08%). The lowest ash content was observed in NERICA-NP (0.52%). However, NERICA-Direct had a significant highest protein content (14.98%) and lowest carbohydrate content (83.66%).

Generally, the parboiled samples had higher content of ash, protein and lower carbohydrate compared to their non-parboiled counterparts.

**Table 1. Proximate composition of TOX 3145 and NERICA-3 varieties (% dry matter)**

Samples	Fat %	Ash %	Protein %	Carbohydrate %
TOX- NP	0.39±0.06 <sup>a</sup>	0.57±0.04 <sup>b</sup>	8.97±0.08 <sup>d</sup>	90.08±0.01 <sup>a</sup>
TOX- IRAD	0.44±0.03 <sup>a</sup>	0.88±0.01 <sup>ab</sup>	10.83±0.11 <sup>c</sup>	87.85±0.06 <sup>b</sup>
TOX-Trad	0.58±0.01 <sup>a</sup>	1.19±0.26 <sup>a</sup>	12.52±0.14 <sup>b</sup>	85.71±0.11 <sup>c</sup>
NERICA-NP	0.51±0.15 <sup>a</sup>	0.52±0.00 <sup>b</sup>	10.35±0.71 <sup>cd</sup>	88.63±0.86 <sup>ab</sup>
NERICA-Indirect	0.51±0.01 <sup>a</sup>	0.83±0.01 <sup>ab</sup>	11.31±0.18 <sup>bc</sup>	87.36±0.19 <sup>bc</sup>
NERICA-Direct	0.60±0.10 <sup>a</sup>	0.76±0.01 <sup>ab</sup>	14.98±0.12 <sup>a</sup>	83.66±0.00 <sup>d</sup>

Mean±SD values with the same superscript letters in a column are not significantly different ( $P>0.05$ )

Table 2 presented varied concentration of phosphorus, potassium, magnesium and iron contents ranging from 0.176 to 0.289 g/100 g; 0.266 to 0.369 g/100 g; 0.173 to 0.281 g/100 g and 0.0 to 0.017 g/100 g, respectively.

Phosphorus and potassium composition in parboiled samples were significantly higher ( $P < 0.05$ ) compared to their non-parboiled counterparts.

**Table 2. Mineral composition of TOX 3145 and NERICA-3 varieties (g/100 g dry matter)**

Samples	Phosphorus	Potassium	Magnesium	Iron
TOX- NP	0.186±0.002 <sup>d</sup>	0.273±0.006 <sup>b</sup>	0.185±0.003 <sup>ab</sup>	0.007±0.000 <sup>ab</sup>
TOX- IRAD	0.237±0.0001 <sup>c</sup>	0.369±0.002 <sup>a</sup>	0.173±0.005 <sup>b</sup>	0.004±0.600 <sup>b</sup>
TOX-Trad	0.274±0.007 <sup>b</sup>	0.341±0.020 <sup>a</sup>	0.281±0.050 <sup>a</sup>	0.017±0.006 <sup>a</sup>
NERICA-NP	0.246±0.005 <sup>c</sup>	0.266±0.006 <sup>b</sup>	0.246±0.001 <sup>ab</sup>	0.001±0.000 <sup>b</sup>
NERICA-Indirect	0.289±0.001 <sup>a</sup>	0.359±0.020 <sup>a</sup>	0.248±0.001 <sup>ab</sup>	0.000±0.000 <sup>b</sup>
NERICA-Direct	0.236±0.003 <sup>c</sup>	0.280±0.007 <sup>b</sup>	0.214 ±0.013 <sup>ab</sup>	0.000±0.000 <sup>b</sup>

*Mean±SD values with the same superscript in a column are not significantly different ( $P > 0.05$ )*

The starch fractions of TOX 3145 and NERICA-3 rice samples are presented in Table 3. The variation in TS among samples (82.19 - 89.44%) was not statistically different ( $P > 0.05$ ). The most predominant fraction of the TS was the DS fraction which varied between 72.22 to 85.63%.

The fraction of RS ranged from 3.81 to 11.07 %. Highest value for RS was observed in TOX-IRAD (11.07%) while the two non-parboiled samples (TOX-NP and NERICA-NP) showed significantly lowest RS values ( $P < 0.05$ ). TOX-NP had the highest DS (85.63%) while NERICA-Direct had lowest DS (72.22%).

The comparison of samples between different parboiling techniques showed that TOX-IRAD had a higher RS (11.07%) with lower DS (76.08%) than its counterpart, TOX-TP (8.82 and 78.67%, respectively). Similar pattern was observed in samples of NERICA-3 variety. NERICA-Direct had higher RS (10.36%) with lower DS (72.22%) compared to NERICA-Indirect (8.35 and 75.95%, respectively).

**Table 3. Starch fractions of TOX 3145 and NERICA-3 varieties (% dry matter)**

Samples	RS	DS	TS
TOX- NP	3.81±0.97 <sup>b</sup>	85.63±6.88 <sup>a</sup>	89.44±6.45 <sup>a</sup>
TOX- IRAD	11.07±3.09 <sup>a</sup>	76.08±5.40 <sup>ab</sup>	87.15±2.39 <sup>a</sup>
TOX-Trad	8.82±2.66 <sup>a</sup>	78.67±8.13 <sup>ab</sup>	87.48±5.57 <sup>a</sup>
NERICA-NP	4.84±0.13 <sup>b</sup>	77.35±2.10 <sup>ab</sup>	82.19±2.24 <sup>a</sup>
NERICA-Indirect	8.35±0.86 <sup>a</sup>	75.95±2.83 <sup>ab</sup>	84.30±3.63 <sup>a</sup>
NERICA-Direct	10.36±0.24 <sup>a</sup>	72.22±3.36 <sup>b</sup>	82.57±3.58 <sup>a</sup>

*Mean±SD values with the same superscript in a column are not significantly different ( $P > 0.05$ )*

*TS= total starch; DS= digestible starch; RS= resistant starch.*

The starch hydrolysis curves from 0 to 120 min for samples are presented in Fig. 1. All the samples portrayed rapid starch hydrolysis at the first 30 min of digestion. The curves for parboiled samples were lower than the non-parboiled samples. TOX-IRAD and NERICA-Direct did not achieve plateau before 90 min hydrolysis.

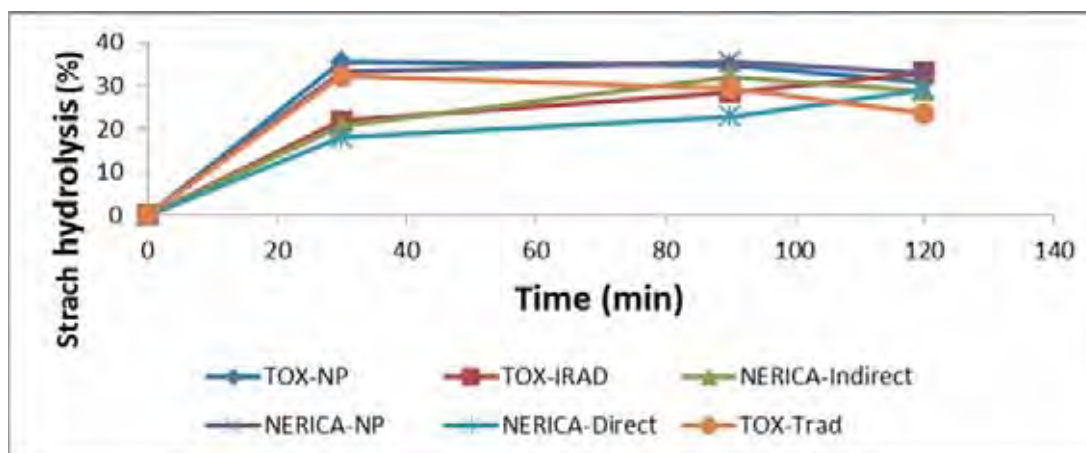


Fig. 1. Starch hydrolysis curves for rice samples (0-120 min).

Table 4 shows varied kinetics of *in vitro* starch digestibility among samples. Extent of starch hydrolysis as indicated by equilibrium concentration ( $C_{\infty}$ ) was highest in the NERICA-NP (34.26%) followed by TOX-NP (33.68%). NERICA-Direct had the lowest  $C_{\infty}$  (27.26%).

The lowest value for kinetic constant ( $k$ ) parameter was observed in NERICA-Direct ( $0.03 \text{ min}^{-1}$ ) followed by NERICA-Indirect and TOX-IRAD ( $0.04 \text{ min}^{-1}$ ). However, NERICA-NP, TOX-NP and TOX-Trad had values of ( $0.2, 0.11, 0.2 \text{ min}^{-1}$ , respectively).

The values for pGI varied from 57.57 to 67.78% among samples. The highest pGI was found in NERICA-NP (67.78%) followed by TOX-NP (66.99%) whereas NERICA-Direct had a significant lowest pGI (57.57%).

Table 4. Kinetics parameters and predicted glycemic index rice samples

Samples	( $C_{\infty}$ )* (%)	( $k$ )* ( $\text{min}^{-1}$ )	$H_{90\text{exp}}$ (%)	pGI	$R^2$
TOX- NP	33.68	0.20	34.60±1.12 <sup>a</sup>	66.99±0.89 <sup>a</sup>	0.98
TOX- IRAD	31.69	0.04	28.59±2.73 <sup>ab</sup>	62.17±2.19 <sup>ab</sup>	0.98
TOX-Trad	28.47	0.20	29.49±2.06 <sup>ab</sup>	62.89±1.65 <sup>ab</sup>	0.93
NERICA-NP	34.26	0.11	35.58±0.13 <sup>a</sup>	67.78±0.10 <sup>a</sup>	0.99
NERICA-Indirect	30.97	0.04	32.06±1.58 <sup>a</sup>	64.96±1.27 <sup>a</sup>	0.98
NERICA-Direct	27.26	0.03	22.86±0.70 <sup>b</sup>	57.57±0.56 <sup>b</sup>	0.96

Mean±SD values within same column followed by same letters are not significantly different ( $P>0.05$ ).

\*Parameters of model equation  $C = C_{\infty} (1 - e^{-kt})$ ;  $C_{\infty}$  = equilibrium concentration;  $k$  = kinetic constant;  $H_{90\text{exp}}$ : experimental values for total starch hydrolysis at 90 min; pGI = predicted glycemic index.

Table 5 presented the relationship of pGI and starch fractions with nutrient composition. RS had significantly positive relationship with protein ( $r = 0.630$ ;  $P < 0.05$ ), ash ( $r = 0.788$ ;  $P < 0.01$ ), fat ( $r = 0.634$ ;  $P < 0.05$ ), phosphorus ( $r = 0.659$ ;  $P < 0.05$ ) and magnesium ( $r = 0.728$ ;  $P < 0.01$ ). On the other hand, RS was inversely related to carbohydrate ( $r = -0.57$ ;  $P > 0.05$ ), DS ( $r = -0.747$ ;  $P < 0.01$ ) and pGI ( $r = -0.143$ ;  $P > 0.05$ ). An inverse but weak correlation was observed between pGI and RS ( $-0.143$ ;  $P > 0.05$ ).

The pGI showed negative correlation with protein, ash, fat, phosphorus, potassium and RS, although the correlation was not significant ( $P > 0.05$ ).

**Table 5. Pearson correlation coefficient (r) between some parameters of starch digestibility and nutrient composition**

Parameters	Ash	CHO	Protein	Fat	P	Mg	K	RS
Ash	1							
CHO	-0.459 <sup>NS</sup>	1						
Protein	0.428 <sup>NS</sup>	-0.951 <sup>**</sup>	1					
Fat	0.383 <sup>NS</sup>	-0.659 <sup>*</sup>	0.769 <sup>*</sup>	1				
P	0.523 <sup>NS</sup>	-0.768 <sup>**</sup>	0.895 <sup>**</sup>	0.777 <sup>**</sup>	1			
Mg	0.264 <sup>NS</sup>	-0.370 <sup>NS</sup>	0.56 <sup>NS</sup>	0.336 <sup>NS</sup>	0.655 <sup>*</sup>	1		
K	0.643 <sup>*</sup>	-0.704 <sup>*</sup>	0.648 <sup>*</sup>	0.351 <sup>NS</sup>	0.605 <sup>*</sup>	0.065 <sup>NS</sup>	1	
RS	0.788 <sup>**</sup>	-0.568 <sup>NS</sup>	0.630 <sup>*</sup>	0.634 <sup>*</sup>	0.659 <sup>*</sup>	0.728 <sup>**</sup>	0.395 <sup>NS</sup>	1
DS	-0.529 <sup>NS</sup>	0.178 <sup>NS</sup>	-0.305 <sup>NS</sup>	-0.320 <sup>NS</sup>	-0.467 <sup>NS</sup>	-0.671 <sup>*</sup>	0.035 <sup>NS</sup>	-0.747 <sup>**</sup>
pGI	-0.293 <sup>NS</sup>	0.400 <sup>NS</sup>	-0.299 <sup>NS</sup>	-0.113 <sup>NS</sup>	-0.212 <sup>NS</sup>	0.110 <sup>NS</sup>	-0.227 <sup>NS</sup>	-0.143 <sup>NS</sup>

Levels of significance: \*\* $P < 0.01$ , \* $P < 0.05$ ; NS: Not significant ( $P > 0.05$ )  
 CHO=carbohydrate; P=Phosphorus; Mg=Magnesium; K=Potassium

#### 4. DISCUSSION

The proximate composition of samples in this study is consistent with a previous report on improved rice varieties [16]. Adu-Kwarteng et al. [16] assessed the nutritional composition of 10 newly improved rice varieties in Ghana. Their study reported protein values between 5.95 to 9.16% and ash content of 0.45-0.59%.

In the case of mineral composition, Adu-Kwarteng et al. [16] reported lower values for phosphorus (0.121-0.199 g/100 g) and potassium (0.071-0.102 g/100 g) compared to values in this present study. The variation could be attributed to geographical location and varietal differences [16].

The significantly lower iron content among the NERICA-3 samples in comparison with samples of TOX variety supports varietal differences as a factor influencing nutritional composition.

Observation of generally poor iron content among samples is not surprising since milled rice is characterized by very low iron content [27].

The finding of higher levels of protein and ash among the parboiled samples as compared to their non-parboiled counterparts is in agreement with literature report [28]. Damir [28] attributed the increased ash content in parboiled rice to penetration of water soluble mineral salt during soaking and steaming process.

This study therefore, supports nutrient solubilisation and consequent higher concentration of nutrient in the endosperm during parboiling [7,8]. This explains the higher losses of nutrients during milling in non-parboiled than parboiled rice [6].

Similarly, the higher phosphorus and potassium among parboiled samples can be explained by their significantly higher protein and ash contents.

This finding however, portrays the relevance of parboiling process on enhanced nutritional value of rice grain particularly on starch digestibility.



Various nutrients including protein, moisture, and phosphorus in food have been associated to starch digestibility [17,19-22]. High protein content in food as well as protein-starch complex were related to low glycemic response [17,18]. Previous studies have demonstrated the influence of high phosphorus content in lowering starch digestibility rate [20,21].

The RS content among studied samples are comparable with literature data on milled rice [29, 30, 25]. The finding of higher RS with a lower DS fraction in parboiled rice than non-parboiled rice is consistent with literature documentation [31-34].

Walter et al. [29] reported a lower range (0.6-5.0%) of RS in non-parboiled samples compared to 2.3-5.8% found in parboiled rice samples. The DS fractions in their study had mean values of 82.4 for non-parboiled rice and 73.9% for parboiled samples.

The tendency of gelatinized starch during parboiling to undergo retrogradation upon cooling leads to resistant starch formation in parboiled rice grain [35].

The variation found in RS fraction between the different parboiled samples could be attributed to application of different degree of heat treatment during the different parboiling process. The protocol of traditional parboiling involved non-uniform distribution of steam during parboiling which could have influenced the sampled grain.

RS fraction has demonstrated a significant influence on glycemic index of rice [30,24,36,25].

The pGI values among samples are within the moderate glycemic index of food classification [37]. This result is consistent with previous reports on milled non-parboiled rice. Hu et al. [30] reported 60.1-106.3% glycemic index in milled rice varieties. Miller et al. [38] showed glycemic index values between 64 and 93%.

This finding of higher pGI among the non-parboiled samples (NERICA-NP and TOX-NP) is in agreement with literature reports that parboiling elicits lower glycemic response compared to the non-parboiled rice [9-12].

Moreover, the lower pGI among parboiled samples supports earlier reports that higher protein and phosphorus content in food reduced rate of starch digestibility [17,20- 22]. The highly significant and positive correlation ( $r=0.895$ ;  $P<0.01$ ) between phosphorus and protein content in our study confirms the influence of protein and phosphorus on starch digestibility.

Deepa et al. [23] reported similar correlation of DS and RS with proximate nutrients in brown rice varieties.

Our study therefore, supports the literature fact that nutritional composition influences starch digestibility. The formation of complex with starch by protein, fat and phosphorus hinders starch susceptibility to amylolytic enzymes action [17,19-23].

The starch kinetic was described by a non-linear model [25] which indicates higher resistance to enzymatic hydrolysis with a lower rate of starch digestion in NERICA-Direct and TOX-IRAD compared to other samples in this study.

## 5. CONCLUSION

The nutritional and starch digestibility profiles of parboiled and non-parboiled rice samples were highlighted. The parboiled samples had higher RS, lower pGI, and kinetic constant compared to non-parboiled samples. RS had a significant positive relationship with nutritional composition, while pGI was inversely correlated to protein, ash, fat, phosphorus, potassium and RS.

This study portrayed the relationship of nutritional composition with starch digestibility in rice. There is need to extend this study on more rice varieties that are processed with similar parboiling techniques.

## ACKNOWLEDGEMENTS

This study was supported by the Canadian International development Agency (CIDA) project on enhancing food security in Africa through the improvement of rice post-harvest handling, marketing and the development of new rice-based products.

## COMPETING INTERESTS

Authors have declared that no competing interests exist.

## REFERENCES

1. Goufo P. Evaluating the constraints and opportunities for sustainable rice production in Cameroon. *Res J Agri Biol Sci.* 2008;4(6):734-744.
2. Wolever TMS, Jenkins DJA, Jenkins AL, Josse RG. The glycemic index: methodology and clinical implications. *Am J Clin Nutr.* 1991;54:846–854.
3. Foster-Powell K, Holt SHA, Brand-Miller JC. International table of glycemic index and glycemic load values: 2002. *Am J Clin Nutr.* 2002;76(1):5-56.
4. Pedersen B, Eggum B. The influence of milling on the nutritive value of flour from cereal grains. 4. Rice. *Plant Foods Hum Nutr.* 1983;33(4):267-278.
5. Bello M, Baeza R, Tolaba M. Quality characteristics of milled and cooked rice affected by hydrothermal treatment. *J Food Eng.* 2006;72(2):124-133.
6. Doesthale YG, Devara S, Rao S, Belavady B. Effect of milling on mineral and trace element composition of raw and parboiled rice. *J Sci Food Agri.* 1979;30(1):40-46.
7. Heinemann R, Fagundes P, Pinto E, Penteado M, Lanfer-Marquez U. Comparative study of nutrient composition of commercial brown, parboiled and milled rice from Brazil. *J Food Comp Anal.* 2005;18(4):287-296.
8. Amato GW, Carvalho JLV, Silveira FS. In: Ricardo Lenz (Ed.), *Arroz parboiled: tecnologia limpa, produto nobre.* Porto Alegre, RS, Brazil, 240pp. World Health Organization, Diet, nutrition and the prevention of chronic diseases. Report of a Joint WHO/FAO Expert Consultation. WHO Technical Report Series 916; 2003.
9. Larsen HN, Rasmussen OW, Rasmussen PH, Alstrup KK, Biswas SK, Tetens I, *et al.* Glycemic index of parboiled rice depends on the severity of processing: study in type 2 diabetic subjects. *Eur J Clin Nutr.* 2000;54:380–385.
10. Jarvi AE, Karlstrom BE, Granfeldt YE, Bjorck I, Vessby B, Asp N. The influence of food structure on postprandial metabolism in patients with non-insulin-dependent diabetes mellitus. *Am J Clin Nutr.* 1995;61(4):837-842.

11. Casiraghi MC, Brighenti F, Pellegrini N, Leopardi E, Testolin G. Effect of processing on rice starch digestibility evaluated by in vivo and in vitro methods. *J Cereal Sci.* 1993;17:147–156.
12. Jenkins DJA, Wolever TMS, Jenkins AL. Starchy foods and glycemic index. *Diabetes Care.* 1988;11(2):149-159.
13. Agama-Acevedo E, Islas-Hernández JJ, Pacheco-Vargas G, Osorio-Díaz P, Bello-Pérez LA. Starch digestibility and glycemic index of cookies partially substituted with unripe banana flour. *Food Sci Technol.* 2012;46(1):177-182.
14. Olembo N, M'mboyi F, Oyugi K. Success Stories in Crop Improvement in Africa; 2010.
15. West Africa Rice Development Association, (WARDA). Annual report. 2002-2003. The Africa Rice Center, Bouake', Cote D'ivoire, <http://www.Warda.Org>. (retrieved, 13<sup>th</sup> August 2007).
16. Adu-Kwarteng E, Ellis W, Oduro I, Manful J. Rice grain quality: a comparison of local varieties with new varieties under study in Ghana. *Food Control.* 2003;14(7):507-514.
17. Chung HJ, Liu Q, Hoover R, Warkentin TD, and Vandenberg B. In vitro starch digestibility, expected glycemic index, and thermal and pasting properties of flours from pea, lentil and chickpea cultivars. *Food Chem.* 2008;111:316-21.
18. Anderson IH, Levine AS, and Levitt MD. Incomplete absorption of the carbohydrate in all-purpose wheat flour. *New England J Med.* 1981;304:891-2.
19. Lynch D, Liu Q, Tarn T, Bizimungu B, Chen Q, Harris P, Chik C, and Skjodt N. Glycemic index — a review and implications for the potato industry. *Am J Potato Res.* 2007;84:179-90.
20. Absar N, Zaidul ISM, Takigawa S, Hashimoto N, Matsuura-Endo C, Yamauchi H, and Noda T. Enzymatic hydrolysis of potato starches containing different amounts of phosphorus. *Food Chem.* 2009;112:57-62.
21. Noda T, Takigawa S, Matsuura-Endo C, Suzuki T, Hashimoto N, Kottearachchi NS, Yamauchi H, Zaidul ISM. Factors affecting the digestibility of raw and gelatinized potato starches. *Food Chem.* 2008;110:465-70.
22. Sitohy MZ, Ramadan MF. Degradability of different phosphorylated starches and thermoplastic films prepared from corn starch phosphomonoesters. *Starch – Starke.* 2001;53:317-322.
23. Deepa G, Singh V, Naidu KA. A comparative study on starch digestibility, glycemic index and resistant starch of pigmented ('Njavara' and 'Jyothi') and non-pigmented ('IR 64') rice varieties. *J Food Sci Technol.* 2010;47(6):644–649.
24. Frei M, Siddhuraju P, Becker K. Studies on the in vitro starch digestibility and the glycemic index of six different indigenous rice cultivars from the Philippines. *Food Chem.* 2003;83(3):395-402.
25. Goñi I, Garcia-Alonso A, Saura-Calixto F. A starch hydrolysis procedure to estimate glycemic index. *Nutr Res.* 1997;17(3):427-37.
26. *Official Methods of Analysis (AOAC)*. Association of Official Analytical Chemist, International. 17th ed. Gaithersburg, Maryland, USA; 2000.
27. Lucca P, Hurrell R, Potrykus I. Fighting iron deficiency anemia with iron-rich rice. *J Am Coll Nutr.* 2002;21(suppl 3):184S-190S.
28. Damir A. Comparative studies on the physicochemical properties and microstructure of raw and parboiled rice. *Food Chem.* 1985;16(1):1-14.
29. Walter M, Da Silva LP, Denardin CC. Rice and resistant starch: different content depending on chosen methodology. *J Food Comp Anal.* 2005;18(4):279-285.
30. Hu P, Zhao H, Duan Z, Linlin Z, Wu D. Starch digestibility and the estimated glycemic score of different types of rice differing in amylose contents. *J Cereal Sci.* 2004;40(3):231-237.

31. Newton J, Wang YJ, Mauromoustakos A. Effects of Cultivar and Processing Condition on Physicochemical Properties and Starch Fractions in Parboiled Rice. *Cereal Chem.* 2011;88(4):414-420.
32. Tetens I, Biswas SK, Glito LV, Kabir KA, Thilsted SH, Choudhury NH. Physicochemical characteristics as indicators of starch availability from milled rice. *J Cereal Sci.* 1997;26:355-361.
33. Eggum B, Juliano B, Perez C, Acedo E. The resistant starch, undigestible energy and undigestible protein contents of raw and cooked milled rice. *J Cereal Sci.* 1993;18(2):159-170.
34. Marsono Y, Topping DL. Complex carbohydrates in Australian rice products—Influence of microwave cooking and food processing. *Lebensm. Wiss. Technol.* 1993;26:364-370.
35. Mitra A, Bhattacharya D, Roy S. Role of resistant starches particularly rice containing resistant starches in type 2 diabetes. *J Hum Ecol.* 2007;21(1):47-51.
36. Rashmi S, Urooj A. Effect of processing on nutritionally important starch fractions in rice varieties. *Inter J Food Sci Nutr.* 2003;54(1):27-36.
37. Jenkins DJ, Kendall CW, Augustin LS, Franceschi S, Hamidi M, Marchie A, Jenkins AL, Axelsen M. Glycemic index: overview of implications in health and disease. *Amer J Clin Nutr.* 2002;76(1):266S-273S.
38. Miller JB, Pang E, Bramall L. Rice: a high or low glycemic index food? *Am J Clin Nutr.* 1992;56:1034–1036.

---

© 2013 Odenigbo et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Peer-review history:*

*The peer review history for this paper can be accessed here:*  
<http://www.sciencedomain.org/review-history.php?iid=266&id=30&aid=2429>

# Study on the gelatinization properties and amylose content of rice varieties from Nigeria and Cameroun

Amaka M. Odenigbo<sup>1,2</sup>, Michael Ngadi<sup>1,\*</sup>, Chijioke Ejebe<sup>1</sup>, Chijioke Nwankpa<sup>1</sup>, Nahemiah Danbaba<sup>3</sup>, Sali Ndindeng<sup>4</sup>, John Manful<sup>5</sup>

<sup>1</sup>Department of Bioresource Engineering, McGill University, 2111 Lakeshore Road, Ste-Anne-de-Bellevue, Quebec, H9X 3V9 Canada

<sup>2</sup>Department of Human Nutrition and Dietetics, Michael Okpara University of Agriculture, Umudike, Nigeria

<sup>3</sup>NCRI, Badeggi, Nigeria

<sup>4</sup>IRAD, Yaounde, Cameroun

<sup>5</sup>AfricaRice, Cotonu, Benin

## Email address:

michael.ngadi@mcgill.ca (M. Ngadi)

## To cite this article:

Amaka M. Odenigbo, Michael Ngadi, Chijioke Ejebe, Chijioke Nwankpa, Nahemiah Danbaba, Sali Ndindeng, John Manful. Study on the Gelatinization Properties and Amylose Content of Rice Varieties from Nigeria and Cameroun. *International Journal of Nutrition and Food Sciences*. Vol. 2, No. 4, 2013, pp. 181-186. doi: 10.11648/j.ijjnfs.20130204.14

**Abstract:** Thirteen varieties (improved and local varieties) of non-parboiled milled rice (*Oryza sativa Linn*) grown in Nigeria and Cameroun were screened for gelatinization and amylose profile. Differential Scanning Calorimeter (DSC) was used in determining the gelatinization enthalpy ( $\Delta H$ ), onset ( $T_o$ ), peak ( $T_p$ ) and conclusion ( $T_c$ ) temperatures. Results from DSC curves presented a single endothermic transition and a flow of maximum heating at peak temperatures from 67.66 and 81.27 °C. The enthalpy levels varied from 0.33 J/g for Panter, to 2.90 J/g for Jamila. Amylose content varied from 8.59 % for FARO 57, to 23.61 % for TOX 3145. Comparing samples of local varieties with those of improved varieties showed higher values for onset and peak gelatinization temperatures among local varieties. A significant and positive correlation was observed in onset temperature, peak and conclusion temperatures while amylose was negatively and weakly related to all gelatinization parameters.

**Keywords:** Non-Parboiled Rice, Improved Rice, Local Rice, Gelatinization, Amylose, DSC

## 1. Introduction

Hybridization between the two domesticated/cultivated rice species; *Oryza sativa Linn* and *Oryza glaberrima Steud* for desirable agronomic characteristics led to the production of New Rice for Africa known as NERICA [1,2]. Presently in Africa, there are several newly improved breeding lines, which are grown along with the previously cultivated varieties [3].

These numerous rice varieties in Africa have wide genetic variation [4].

The gene trait inherited from parental specie has been associated to amylose content in rice varieties. Kishine *et al.* [5] reported that NERICA varieties with gene inheritance from *O. glaberrima* had high amylose content while the varieties with gene from inheritance from the *O. sativa lines* had lower amylose content. Rice classification based on amylose content was described by Lawal *et al.* [6] as follows; waxy (0-2 % amylose),

very low amylose (2 – 12 %), low amylose (12 – 20 %), intermediate amylose (20-25 %) and high amylose (25-33 %).

Amylose and amylopectin are two glucose polymers in starch granules [6]. Amylose is essentially linear, consist of  $\alpha$ -(1,4)-linked D-glucopyranosyl units while amylopectin is highly branched and made up of  $\alpha$ -(1,4)-linked D-glucopyranosyl units joined through  $\alpha$ -(1,6) linkages [7].

Approximately, 90% of milled rice (dry matter) comprised of starch and the eating as well as cooking quality of rice is influenced by the starch characteristics, which include amylose content, gelatinization temperatures [8].

Quantifying gelatinization characteristics of food is very relevant in food processing because it allows simulation of the cooking process for improved functional properties [8,9].

Amylose tends to act as a restraint to gelatinization because it diffuses out of the granules during swelling, making up the continuous phase (network) outside the



granules [10]. It has been reported that waxy starches usually swell to a greater extent than non-waxy starches [11, 12].

Varavinit *et al.* [11] reported a positive correlation of amylose to onset, peak and conclusion temperatures (0.84, 0.88, and 0.85, respectively). Their study suggested the prediction of amylose content of rice with gelatinization properties. A similar result by Park *et al.* [13] stated that gelatinization temperatures increase with higher amylose content in rice starches. Park *et al.* [13] postulated that the crystalline regions of non-waxy (high amylose) rice starch restricted the hydration of amorphous regions whereas waxy (high amylopectin) rice starch consisted mostly of crystalline regions and thus could begin gelatinization at a lower temperature.

Differential Scanning Calorimeter (DSC) has been used extensively to describe the gelatinization properties of rice flour. Normand and Marshall [14] demonstrated only one endothermic transition during gelatinization milled rice flour whereas milled whole grain rice exhibited two endothermic transitions.

This study therefore, aimed at using DSC in screening the gelatinization properties of selected improved and local rice varieties among the popular non-parboiled, milled rice in Nigeria and Cameroun.

## 2. Materials and Methods

### 2.1. Samples

Rice varieties are *Japonica* subspecies, crossed between two *Oryza sativa* parents. The improved varieties are improved lines of *Oryza sativa L* originated from AfricaRice Center, Benin. Samples from Nigeria were five improved varieties (*FARO 44*, *FARO 57*, *FARO 60*, *FARO 61*) obtained from the Breeding Unit of Rice Research Program, National Cereal Research Institute (NCRI) Badeggi and four local varieties (*Kwandala*, *Yardass*, *Jeep*, *Jamila*) obtained from crop improvement unit of Kano State Agriculture and Rural Development Agency (KNARDA). From Cameroun, two improved varieties (TOX 3145, NERICA-3) and two local varieties (*Panter*, *Tianan*) were collected from a rice farmer in Ndop, Northwest Region of Cameroun.

Panicle picking was used in harvesting the samples and afterward samples were dried to 12 % moisture content before manually threshed and cleaned. The samples were dehulled using a rice dehuller (THU 35A, Satake, Engineering Co. Ltd., Tokyo) before milling (McGill mill, Model No. 2, Brookshire, Texas) for 30 seconds.

### 2.2. Sample Preparation for Analysis

Grains of the rice samples were individually ground using a coffee grinder (SUMEET Multi Grind, India) and passed through a 0.5 mm sieve (CETyler, Ontario, Canada).

### 2.2.1. Isolation of Starch

Starch was extracted from rice flour samples by alkaline deproteination method according to Lim *et al.* [15] with slight modification. About 100g of flour was mixed with 300 ml of 0.5 % NaOH. The mixture was constantly stirred for 4 h and left to stand for 24 h at 10 °C. The supernatant was decanted and the solid phase washed several times with distilled water until the pH of the filtrate was between 6.0 and 6.5. The isolated starch was oven-dried at 40 °C for 48 h and later ground in a mortar to pass through a 0.55 mm mesh sieve, then used for amylose determination.

### 2.3. Gelatinization

The gelatinization properties were examined by a Differential Scanning Calorimeter (DSC Q100, TA instruments, Wilmington, DE, USA). Heat flow and temperature calibrations of the DSC were performed using pure indium with 28.41 J/g heat of fusion and a melting temperature of 156.66 °C. The experimental values were within  $\pm 3$  % of the literature value.

A mass of  $3 \pm 0.01$  mg was carefully placed at the centre of a pre weighed aluminium pan (40  $\mu$ l). Considering the moisture content of each sample, appropriate volume of distilled water was added to the pan by a micropipette to achieve a flour/ water ratio of 1: 2. This ratio of 1:2 accounts for more than 60 % moisture, which is the moisture content required for rice gelatinization and to obtain a single endotherm during DSC experiment [16]. The pans were hermetically sealed with TA sample encapsulating press. The sealed pans were stored at room temperature to stabilize for one hour before analysis. The sample pans were placed in the sample cells while an empty pan was placed in the reference cell of the DSC. Samples were heated from 35 to 100 °C at a rate of 10 °C/min. The onset ( $T_o$ ), peak ( $T_p$ ) and conclusion ( $T_c$ ) temperatures of gelatinization were determined. Gelatinization enthalpy ( $\Delta H$ ) was measured in J/g of dried sample weight.

### 2.4. Amylose Determination

The analytical procedure of Hoover and Ratnayake [17] was adopted in determination of amylose content. Isolated starch (20 mg) derived from procedure in 2.2.1 was vortexed with 8 ml of 90% dimethylsulfoxide (DMSO) then incubated in a shaking water bath at 85 °C for 15 min. Solution was subsequently cooled at room temperature for 45 min then diluted with distilled and deionized water (DDW) to 25 ml. Aliquot of 1ml from the diluted solution was added to 40 ml of DDW then mixed vigorously with 5 ml of iodine solution (0.0025 M I<sub>2</sub>/0.0065 M KI mixture). Afterward volume was adjusted to 50 ml with DDW, mixed vigorously and allow to develop colour for 15 min. Absorbance was measured at 600 nm against a reagent blank. The percentage of amylose was calculated from an

equation obtained from the standard curve.

### 2.5. Statistical Analysis

Data were subjected to one way analysis of variance (ANOVA) followed by Fisher's least significant-difference (LSD) test ( $P < 0.05$ ) to determine the differences in amylose contents and gelatinization parameters among the thirteen samples. Results were expressed as mean  $\pm$  standard deviation of three replicates. Independent T-test was performed for differences in amylose contents and gelatinization parameters between local and improved rice varieties. The statistical software used was SAS version 9.3 (SAS Institute Inc., Cary, NC, USA). A dendrogram was drawn using UnscramblerX version 10.3 (CAMO Software AS, Oslo, Norway) for cluster analysis.

## 3. Results and Discussion

Gelatinization properties of improved and local rice varieties obtained from DSC curves presented single endothermic transition (Fig. 1 and 2, respectively). The displayed endotherm curves confirmed an earlier report that milled rice flour exhibit only one endothermic transition during gelatinization [14].

Gelatinization parameters and amylose content showed significant variation among samples (Table 1). The gelatinization temperatures ranged from 63.42 to 78.34°C for onset temperature; 67.66 to 81.27°C for peak temperature; and from 72.03 to 90.18 °C for conclusion temperature. This result is similar to those reported by previous studies [4, 14, 18]. The study on 39 milled rice samples from West Africa reported ranges for onset temperature from 58 to 72 °C; peak from 66 to 77 °C and conclusion temperature from 70 to 83 °C [4].

Highest peak temperature (81.27 °C) was found in *FARO 57* which also had the lowest amylose (8.59 %). This implies high content of amylopectin in *FARO 57* sample in this study. This result agrees with the finding of highest peak temperature among waxy rice samples in an earlier

study by Zhu *et al.* [19]. Peak temperature represents half the conversion temperature of sample melting.

The lowest onset and peak temperatures with highest amylose content (23.61 %) was found in TOX 3145. This finding is contrary to the positive correlation of amylose to onset, peak and conclusion temperatures reported by Park *et al.* [11, 13]. This finding suggests that other factors such as starch structures, nutritional composition could have influenced the gelatinization of non-parboiled rice

The low enthalpy from 0.33 to 2.90 J/g found among samples in this present study is comparable to the enthalpy data (from 0.58 to 4.21 J/g) reported on rice varieties from West Africa [4]. Fan *et al.* [18] reported that non-starch components in rice flour such as protein, ash, fiber and lipids reduce enthalpy for gelatinization. Saif *et al.* [20] stated that starch concentration and varied level of water / starch ratio have significant influence on gelatinization properties of rice flour.

Low enthalpy values were also attributed to low molecular weight and chain length distribution of amylopectin [21]. This indicates the need for further study on starch structure of these non-parboiled rice samples in our study.

According to food classification based on amylose content [6], only *FARO 57* and TOX 3145 had very low amylose (8.59 %) and intermediate amylose (23.61 %), respectively. The other samples showed low amylose content (12 to <20 %). The observed levels of amylose are consistent with data in previous studies on milled rice varieties [22, 23]. Almost all samples studied by Bocevskaa *et al.* [22] had low amylose content (12–19.4 %) with exception of one variety that had intermediate amylose content (23.3 %). Similarly, amylose values between 2.3 % and 15.4 % were found among milled rice varieties involved in a study by Singh *et al.* [23].

The variation in amylose classification among samples between very low to intermediate agrees with literature that amylose content in rice is influenced by variety [24, 25].

**Table 1.** Gelatinization properties and amylose content of non-parboiled rice varieties

Rice varieties	To (°C)	Tp (°C)	Tc (°C)	$\Delta H$ (J/g)	Amylose (%)
Improved					
FARO 44	63.50 $\pm$ 2.06f	69.28 $\pm$ 1.48e	79.66 $\pm$ 0.78f	2.51 $\pm$ 0.32ab	16.54 $\pm$ 1.47bc
FARO 52	63.74 $\pm$ 0.55ef	68.79 $\pm$ 0.28e	77.70 $\pm$ 1.10f	1.54 $\pm$ 0.74cde	15.86 $\pm$ 1.16bc
FARO 57	76.47 $\pm$ 0.51b	81.27 $\pm$ 0.44a	89.55 $\pm$ 0.47ab	2.13 $\pm$ 0.06bc	8.59 $\pm$ 1.24d
FARO 60	76.97 $\pm$ 0.17ab	80.31 $\pm$ 0.12ab	90.18 $\pm$ 2.74a	1.17 $\pm$ 0.01ef	18.49 $\pm$ 2.01ab
FARO 61	78.34 $\pm$ 0.37a	81.22 $\pm$ 0.35a	87.86 $\pm$ 0.91abc	1.28 $\pm$ 0.23def	12.96 $\pm$ 2.32bcd
TOX 3145	63.42 $\pm$ 0.04f	67.66 $\pm$ 0.16f	74.02 $\pm$ 1.14g	0.66 $\pm$ 0.16fg	23.61 $\pm$ 2.32a
NERICA- 3	73.76 $\pm$ 0.36c	76.90 $\pm$ 0.43c	83.56 $\pm$ 0.47e	0.62 $\pm$ 0.12fg	12.46 $\pm$ 0.17cd
Local					
Kwandala	75.88 $\pm$ 0.06b	79.31 $\pm$ 0.02b	87.87 $\pm$ 0.49abc	2.00 $\pm$ 0.26bcd	18.57 $\pm$ 4.34ab
Yardass	76.89 $\pm$ 0.73ab	80.81 $\pm$ 0.30a	87.80 $\pm$ 0.28bc	2.34 $\pm$ 0.43ab	16.26 $\pm$ 3.77bc

Jeep	77.46±0.21ab	80.75±0.16a	86.87±0.51cd	1.38±0.13def	13.39±3.05bcd
Jamila	65.72±0.47d	70.34±0.85d	79.66±0.23f	2.90±0.78a	16.71±3.38bc
Panter	76.84±0.37ab	80.00±0.28ab	85.47±0.64ed	0.33±0.08g	11.74±1.07cd
Tianan	65.21±0.30ed	69.86±0.30ed	72.03±0.39g	1.21±0.09ef	13.68±0.97bcd

Mean± SD values within same column followed by same letters are not significantly different ( $P>0.05$ )

**Table 2.** Comparison of gelatinization parameters and amylose content between improved and local varieties

Parameters	Local	Improved	T-value	P-value
To (°C)	73.00±5.60	70.88±6.75	-0.86	0.09
Tp (°C)	76.84±5.02	75.06±6.03	-0.81	0.14
Tc (°C)	83.28±6.02	83.21±6.13	-0.03	0.81
ΔH (J/g)	1.69±0.92	1.42±0.72	-0.86	0.53
Amylose (%)	15.06±3.31	15.50±4.79	0.27	0.30

$P>0.05$  is not significant

The comparison between improved and local varieties in gelatinization properties and amylose content is presented in Table 2. Local varieties showed higher values for onset temperature (73.00 vs. 70.88 °C), peak temperatures (76.48 vs. 74.06 °C), conclusion temperature (83.23 vs. 83.21 °C) and enthalpy (1.69 vs. 1.42 J/g) than the improved varieties. The mean value for amylose content among improved varieties (15.50 %) was higher compared to local varieties (15.06 %). The P-values presented in Table 2 indicated statistical similarity in parameters between the two groups of rice varieties. However, observation of variation in values for thermal properties and amylose content between improved and local varieties was very weak and insignificant ( $P>0.05$ ). This finding indicated similarity among these rice varieties in thermal properties and in amylose content.

**Table 3.** Correlation Coefficient (*r*) between gelatinization parameters and amylose content of samples.

Parameter	To (°C)	Tp (°C)	Tc (°C)	ΔH (J/g)	Amylose (%)
To (°C)	1				
Tp (°C)	0.994*	1			
Tc (°C)	0.902*	0.923*	1		
ΔH (J/g)	-0.183	-0.114	-0.097	1	
Amylose (%)	-0.215	-0.226	-0.206	-0.080	1

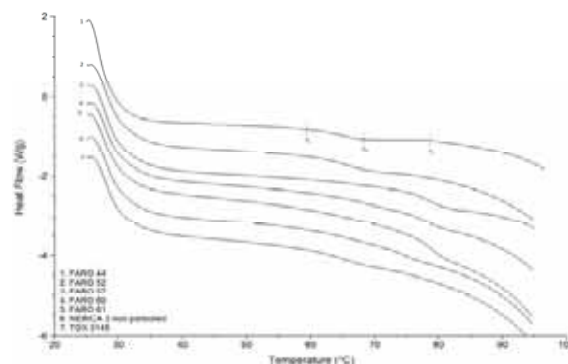
\*\* Correlation is significant at 0.01 level.

The relationship between gelatinization parameters and amylose content is shown in Table 3. The onset, peak and conclusion temperatures were highly and positively correlated ( $P<0.01$ ). The correlation between peak temperature with onset and conclusion temperatures were  $r=0.994$  and  $0.923$ , respectively. The enthalpy and amylose content had inverse and weak correlation with

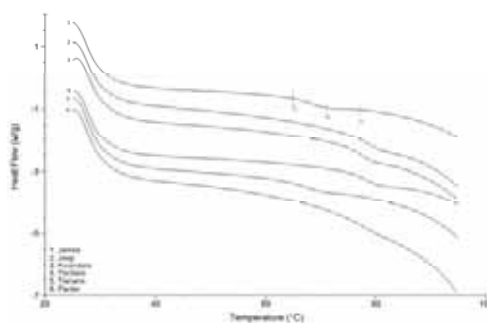
gelatinization temperatures ( $P>0.05$ ). However, positive relationship of amylose with gelatinization properties had been described previously [10, 11, 13, 26].

Multivariate analysis was applied by average linkage clustering to establish similarities in gelatinization properties among samples. Figure 3 presents the similarities among the 13 samples into four major clusters. The first and third clusters comprised only one variety (TOX 3145 and NERICA-3, respectively). The second group contained four varieties (Tianan, Jamila, FARO 52 and FARO 44) while fourth cluster comprised seven varieties (Yardass, Kwandala, FARO 60, FARO 57, Panter, Jeep, FARO 61). TOX 3145 and NERICA-3 indicated diverse thermal characteristics from the other rice samples.

The distinct characteristic of TOX 3145 from the other samples could be explained by its low onset, peak, conclusion temperatures, enthalpy and highest amylose content.



**Figure 1.** The gelatinization endotherm curves of seven improved varieties of non-parboiled rice as determined by differential scanning calorimetry (DSC).



**Figure 2.** The gelatinization endotherm curves of six local varieties of non-parboiled rice as determined by differential scanning calorimetry (DSC).

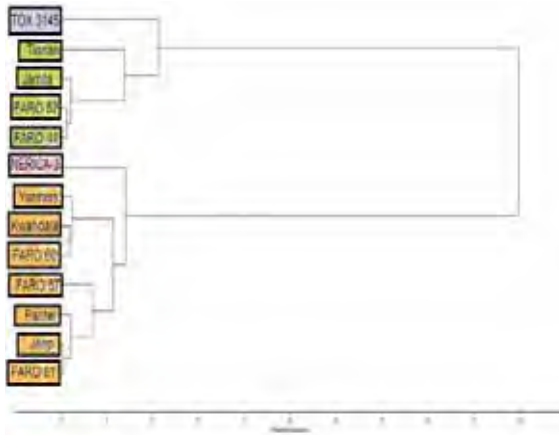


Figure 3. Dendrogram based on gelatinization characteristics and amylose content.

#### 4. Conclusion

This study provides information on gelatinization properties and amylose contents of rice samples from Cameroun and Nigeria. The local varieties had higher onset and peak gelatinization temperatures than the improved varieties. The enthalpy levels were generally low among samples. A significant and positive correlation was observed among onset temperature, peak and conclusion temperatures while amylose was inversely related to all gelatinization parameters.

There is need for further study on influence of factors such as starch structures, nutritional composition on the gelatinization properties of these milled rice.

#### References

- [1] Olembo, N., M'mboyi, F. and Oyugi, K. (2010). Success Stories in Crop Improvement in Africa: The Case of Rice in Sub-Saharan Africa. African Biotechnology Stakeholders Forum (ABSF), Nairobi, Kenya.
- [2] WARDA: West Africa Rice Development Association (2004). Annual report. 2002-2003. The Africa Rice Center, Bouake', Cote D'ivoire, <http://www.Warda.Org>. (retrieved, 13<sup>th</sup> August 2007).
- [3] Fofana, M., Futakuchi, K., Manful, J., Bokossa, Y.I., Dossou, J. and Bleoussi, R. (2011). Rice grain quality: A comparison of imported varieties, local varieties with new varieties adopted in Benin. *Food control*, 22(12), 1821-1825.
- [4] Traore, K., McClung, A.M., Fjellstrom, R. and Futakuchi, K. (2011). Diversity in grain physico-chemical characteristics of West African rice, including NERICA genotypes, as compared to cultivars from the United States of America. *International Research Journal of Agricultural Science and Soil Science*, 1(10), 435-448.
- [5] Kishine, M., Suzuki, K., Nakamura, S. and Ohtsubo, K.I. (2008). Grain qualities and their genetic derivation of 7 new rice for Africa (NERICA) varieties. *Journal of agricultural and food chemistry*, 56(12), 4605-4610.
- [6] Lawal, O.S., Lapasin, R., Bellich, B., Olayiwola, T.O., Cesaro, A., Yoshimura, M. and Nishinari, K. (2011). Rheology and functional properties of starches isolated from five improved rice varieties from West Africa. *Food Hydrocolloids*, 25 (7), 1785-1792.
- [7] Delcour, J.A., Bruneel, C., Derde, L.J., Gomand, S.V., Pareyt, B., Putseys, J.A., Wilderjans, E. and Lamberts, L. (2010). Fate of starch in food processing: from raw materials to final food products. *Food Science and Technology*, 1, 87-111.
- [8] Bao, JS, Sun, M, Zhu, LH and Corke, H. (2004). Analysis of quantitative trait loci for some starch properties of rice (*Oryzasativa* L.): thermal properties, gel texture and swelling volume. *Journal of Cereal Science*, 39 (3), 379-385.
- [9] Tribess, T., Hernández-Urbe, J., Méndez-Montealvo, M., Menezes, E., Bello-Perez, L. and Tadini, C. (2009). Thermal properties and resistant starch content of green banana flour (*Musa cavendishii*) produced at different drying conditions. *LWT-Food Science and Technology*, 42(5), 1022-1025.
- [10] Hermansson, M. and Svegmarm, K. (1996). Developments in the understanding of starch functionality. *Trends in Food Science and Technology*, 7, 345-353.
- [11] Varavinit, S., Shobsngob, S., Varayanond, W., Chinachoti, P. and Naivikul, O. (2003). Effect of amylose content on gelatinization, retrogradation and pasting properties of flours from different cultivars of Thai rice. *Starch-Stärke*, 55(9), 410-415.
- [12] Tester, R. F. and Morrison, W. R. (1990). Swelling and gelatinization of cereal starches. Effects of amylopectin, amylose and lipids. *Cereal Chemistry*, 67, 551-557.
- [13] Park, I.M., Ibáñez, A.M., Zhong, F. and Shoemaker, C.F. (2007). Gelatinization and Pasting Properties of Waxy and Non-waxy Rice Starches. *Starch-Stärke*, 59(8), 388-396.
- [14] Normand, F. L., and W. E. Marshall. (1989). Differential scanning calorimetry of whole milled rice and milled rice flour. *Cereal Chemistry*, 66 (4), 317-320.
- [15] Lim, S.T., Lee, J.H., Shin, D.H. and Lim, H.S. (1999). Comparison of protein extraction solutions for rice starch isolation and effects of residual protein content on starch pasting properties. *Starch-Stärke*, 51, 410-415.
- [16] Billiaderis, C.G., Page, C.M., Maurice, T.J., and Juliano, B.O. 1986. Thermal characterization of rice starch: A polymeric approach to phase transition of granular starch, *Journal of Agricultural and Food Chemistry*, 34, 6-14.
- [17] Hoover, R. and Ratnayake, W. (2001). Determination of total amylose content of starch. *Current protocols in food analytical chemistry*, Wiley, New York (2001). Section E, Unit 2-3.
- [18] Fan, J., Marks, B. P., Daniels, M. J. and Siebenmorgen, T. J. (1999). Effects of postharvest operations on the gelatinization and retrogradation properties of long-grain rice. *Transactions of the ASABE*, 42(3), 727-731.
- [19] Zhu, L. J., Liu, Q. Q., Wilson, J. D., Gu, M. H. and Shi, Y. C. (2011). "Digestibility and physicochemical properties of rice (*Oryza sativa* L.) flours and starches differing in amylose

- content." *Carbohydrate Polymers*, 86(4),1751-1759.
- [20] Saif, S., Lan, Y. and Sweat, V. (2003). Gelatinization properties of rice flour. *International Journal of Food Properties*, 6(3), 531-542.
- [21] Jayakody, L., Hoover, R., Liu, Q. and Donner, E. (2007). Studies on tuber starches. II. Molecular structure, composition and physicochemical properties of yam (*Dioscorea sp.*) starched grown in Sri Lanka. *Carbohydrate polymers*, 69(1), 148 – 163.
- [22] Bocevaska, M., Aldabas, I., Andreevska, D. and Ilieva, V. (2009). Gelatinization behavior of grains and flour in relation to physico-chemical properties of milled rice (*oryza sativa l.*). *Journal of Food Quality*, 32(1), 108-124.
- [23] Singh, J., Dartois, A. and Kaur, L. (2010). Starch digestibility in food matrix: a review. *Trends in Food Science and Technology*, 21(4), 168-180.
- [24] Vlachos, A. and Arvanitoyannis, I.S. (2008). A review of rice authenticity/adulteration methods and results. *Critical reviews in Food Science and Nutrition*, 48(6), 553-598.
- [25] Frei, M., Siddhuraju, P. and Becker, K. (2003). Studies on the in vitro starch digestibility and the glycemic index of six different indigenous rice cultivars from the Philippines. *Food Chemistry*, 83(3), 395-402.
- [26] Jane, J., Chen, Y., Lee, L., McPherson, A., Wong, K., Radosavljevic, M. and Kasemsuwan, T. (1999). Effects of Amylopectin Branch Chain Length and Amylose Content on the Gelatinization and Pasting Properties of Starch 1. *Cereal Chemistry*, 76(5), 629-637.



Original article

## Starch fraction profiles of milled, nonparboiled rice varieties from Nigeria

Amaka M. Odenigbo,<sup>1,2</sup> Michael Ngadi,<sup>1,\*</sup> John Manful<sup>3</sup> & Nahemiah Danbaba<sup>4</sup>

1 Department of Bioresource Engineering, McGill University, 21111 Lakeshore Road, Ste-Anne-de-Bellevue, QC, Canada H9X 3V9

2 Michael Okpara University of Agriculture, Umudike, P.M.B. 7267 Umuahia, Abia state, Nigeria

3 Africa Rice Center (AfricaRice), 01 BP 2031 Cotonou, Benin

4 National Cereals Research Institute (NCRI), Badeggi, P.M.B. 8 Bida, Niger state, Nigeria

(Received 28 November 2012; Accepted in revised form 8 June 2013)

**Summary** This study determined the levels of nutritionally important starch fractions in selected milled, nonparboiled rice cultivated in Nigeria. Five improved varieties (*FARO 52*, *FARO 57*, *FARO 44*, *FARO 60* and *FARO 61*) and four local varieties (*Kwandala*, *Yardass*, *Jeep* and *Jamila*) were evaluated. There were significant differences in starch fractions among varieties. Resistant starch (RS) ranged between 1.43% and 3.13%. Rapidly digestible starch (RDS) was lowest in *Jamila* (27.70%) and highest in *FARO 61* (39.26%). Generally, the local varieties had significant higher RS (2.71%) with a lower RDS (32.82%) compared with improved varieties (RS; 1.88% and RDS; 36.07%). RS was inversely related to RDS and starch digestion index (SDI). The SDI had a highly significant positive correlation with RDS ( $r = 0.879$ ,  $P < 0.01$ ). These results highlight the need for further work in the identification of milled, nonparboiled rice varieties with less rapid digestion for its associated health benefits to consumers.

**Keywords** Milled, nonparboiled, rapidly digestible starch, resistant starch, rice, slowly digestible starch.

### Introduction

Rice is an important staple cereal and a good dietary source of starch. The milled rice grain is predominantly comprised of starch (Khatoun & Prakash, 2006). Nutritionally, starch is classified based on its rate of digestion into rapidly digestible starch (RDS), slowly digestible starch (SDS) and resistant starch (Englyst & Hudson, 1996). Resistant starch (RS) is an important measure that indicates starch digestibility. It has been defined as the starch fraction, which escapes enzymatic digestion in the small intestine but passes to the large bowel (Haralampu, 2000). Rapidly digestible starch is the fraction of starch with rapid digestion and causes a rise in glycaemic response while slowly digestible starch refers to the starch fraction that is digested slowly after the RDS (Englyst *et al.*, 1992; Dona *et al.*, 2010). RDS and SDS have been defined as the concentration of *in vitro* starch digested in the first 30 and 120 min, respectively (Rosin *et al.*, 2002). However, numerous protocols are available for *in vitro* starch digestion in foods (Woolnough *et al.*, 2008).

There was variation in starch fractions among rice varieties due to the processing methods of rough rice

(paddy) which include parboiling. Several authors have established that parboiled rice has more RS content than nonparboiled rice (Eggum *et al.*, 1993; Marsono & Topping, 1993; Tetens *et al.*, 1997; Newton *et al.*, 2011). Also, cooking method has been reported as a factor influencing starch fractions in rice (Rashmi & Urooj, 2003). The study of Rashmi & Urooj (2003) demonstrated significant higher RS fraction (% fresh basis) in pressure cooking and steaming (7.6% and 7.3%, respectively) than in boiling (3.71%) and straining (2.62%) methods of cooking rice. RDS fraction was significantly higher in pressure cooking (23.18%) and boiling (22.57%) but lower in steaming method (10.45%). Steaming method had higher SDS fraction (16.21%).

A considerable amount of research has shown that raw samples contain higher levels of RS compared with cooked samples. Han *et al.* (2008) reported greater RS content in uncooked flour of rice than in cooked rice grains. Raw food products store their starch in compact granules, which are poorly digestible (Allen *et al.*, 2012). Cooking of the food products generates heat that breaks down starch granules for more readily hydrolysis by pancreatic  $\alpha$ -amylase, which consequently increases the glycaemic index of cooked food materials. RS content can significantly be

\*Correspondent: E-mail: michael.ngadi@mcgill.ca

reduced during cooking by disrupting crystalline starch structure (Sajilata *et al.*, 2006; Roopa & Premavalli, 2008).

Different absorption rates between resistant starch (RS) and digestible starch (DS) denote their differential metabolic responses. The health benefits of slowly digestible starch and resistant starch in metabolic disorders such as diabetes mellitus and cardiovascular disease have been reported in many nutritional studies (Jenkins *et al.*, 2002; Brand-Miller *et al.*, 2003; Higgins, 2004).

The dietary starch in polished rice was described by Rosin *et al.* (2002). Samples in their study were characterised on dry basis as RS: 2.41%, DS: 84.95%, TS: 87.36%, RDS: 34.10% and SDS: 41.46%. In Nigeria, over sixty new rice varieties have been developed and released for commercial production, but some indigenous farmers' varieties are also being cultivated side by side the improved varieties. These varieties are either parboiled or milled raw before utilisation in diverse rice-based recipes. Hence, this study aimed at evaluating the starch profile of some commonly consumed milled, nonparboiled rice varieties in Nigeria.

## Materials and methods

### Samples

All the nine rice varieties are *Japonica* subspecies, crossed between two *Oryza sativa* parents. The *FARO* varieties are referred to as 'improved' because they are developed through scientific breeding to meet specific production constraints such as increased yield, resistance to diseases and officially released to rice producers. They are medium maturity (120–130 days from planting to harvesting), short plant type and fertiliser responsive varieties. Five improved rice varieties, namely *FARO 52*, *FARO 57*, *FARO 44*, *FARO 60* and *FARO 61*, and four local varieties, namely *Kwandala*, *Yardass*, *Jeep* and *Jamila*, were randomly selected among most popular varieties in cultivating communities in Nigeria. The improved varieties were obtained from the Breeding Unit of Rice Research Program, National Cereal Research Institute (NCRI) Badeggi, Nigeria, while local varieties were obtained from crop improvement unit of Kano State Agriculture and Rural Development Agency (KNARDA). These samples were collected during the 2010 harvest season (December). The samples were harvested by panicle picking and afterwards dried to 14% moisture content before manually threshed and cleaned. About 2 kg of each variety was packaged in a paper bag and transported to NCRI for dehulling and milling.

The samples were dehulled using a dehuller (THU 35A, Satake, Engineering Co. Ltd., Tokyo) before milling (McGill mill, Model No. 2, Brookshire, Texas) for

30 s. The milled and nonparboiled rice grains were then brought to McGill University, Canada for analysis.

### Sample preparation for analysis

#### Uncooked samples

Samples of milled and nonparboiled rice grains were ground using a coffee grinder (SUMEET Multi Grind, Mumbai, India) and passed through a 0.5 mm sieve (CE Tyler, St. Catharines, ON, Canada). The ground samples were stored in air tight plastic bags for analysis. Each sample was analysed in four replicates.

#### Cooked samples

Whole grains of milled and nonparboiled rice ( $50 \pm 0.05$  mg for predicted glycaemic index [pGI], SDS, RDS while  $100 \pm 0.05$  mg for RS, DS, total starch) were weighed into capped tube and boiled in distilled water (5 mL) for 30 min. Samples were homogenised in same tube with cooking water for 1 min using controlled speed (level 2) Tissue-Tearor homogenizer (Biospec Products, Inc. Bartlesville, OK, USA). The *in vitro* starch digestibility experiment commenced immediately after homogenisation.

### Sample analysis

#### Resistant starch

Resistant starch was determined by Megazyme Resistant Starch Assay procedure (AOAC Method 2002.02; AACC Approved Method 32–40) with a slight modification. Boiled and homogenised samples were incubated with 10 mL of HCl–KCl buffer (pH 1.5) and 20 mg pepsin for 1 h at 37 °C. Afterwards, samples were incubated with pancreatic  $\alpha$ -amylase ( $10 \text{ mg mL}^{-1}$ ) solution containing amyloglucosidase (AMG) for 16 h at 37 °C with constant shaking for starch hydrolysis. After hydrolysis, samples were washed thrice with ethanol (99% v/v and 50% ethanol). The separated pellet from supernatant was further digested with 2 M KOH. Digested pellet and supernatant were separately incubated with AMG. Glucose released was measured using a glucose oxidase–peroxidase kit (K-GLOX, Megazyme Bray, Co. Wicklow, Ireland). The absorbance was measured at 510 nm wavelength against the reagent blank. The glucose content of the supernatant and digested pellet was used in calculation of digestible starch (DS) and resistant starch (RS), respectively, by applying the factor of 0.9. Total starch (TS) was then derived as the sum of DS and RS.

#### Rapidly digestible starch (RDS) and slowly digestible starch (SDS)

A modified *in vitro* method based on the procedure of Goni *et al.* (1997) was adopted. The boiled and

homogenised samples were incubated with 10 mL HCl-KCl buffer (pH 1.5) and 20 mg pepsin at 37 °C for 1 h with constant shaking. The pH was raised with the addition of 200 µL pancreatic α-amylase solution (1.5 mg 10 mL<sup>-1</sup> phosphate buffer) and incubated at 40 °C for 45 min. The enzyme reaction was stopped with 70 µL Na<sub>2</sub>CO<sub>3</sub> solution and samples diluted to 25 mL with tris-maleate buffer (pH 6.9). Five (5) mL of pancreatic α-amylase solution (3 U 5 mL<sup>-1</sup> tris-maleate buffer) was then added to the sample and incubated at 37 °C with constant shaking. Aliquots (duplicate) of 1 mL were taken at 30 and 120 min from the samples and placed into boiling water with vigorous shaking for 5 min to inactivate the enzyme reaction.

Aliquots were treated with 3 mL of 0.4 M sodium acetate buffer (pH 4.75) and 60 µL of AMG (3300 U mL<sup>-1</sup>) then incubated at 60 °C for 45 min with constant shaking. The glucose released was measured using a glucose oxidase-peroxidase (GOPOD) kit (K-GLOX, Megazyme Bray, Co. Wicklow, Ireland). Absorbance was measured at 510 nm wavelength against the reagent blank. Glucose was converted into starch by applying the factor of 0.9. The 30 and 120 min hydrolysis represented the rapidly digestible starch (RDS) and slowly digestible starch (SDS), respectively (Rosin *et al.*, 2002).

### Starch digestion index

The equation of Rashmi & Urooj (2003) was adopted for calculation of starch digestion Index (SDI).

$$\text{SDI} = \text{RDS}/\text{TS} \times 100.$$

### Statistical analysis

Data were expressed as mean values of four replicate measurements. Variation in levels of starch fractions among varieties was determined by a one-way analysis of variance (ANOVA) followed by Duncan's multiple range test ( $P < 0.05$ ). Independent *t*-test was performed for comparison of starch fractions and SDI between local and new rice varieties. Relationships between study parameters were calculated with Pearson's correlation coefficients ( $r$ ). Statistical software used was SAS version 4.3 (SAS Institute Inc., 2010, Cary, NC, USA).

### Results and discussion

Table 1 presents the three nutritionally important starch fractions of rice samples on dry basis. The values for RS fraction varied among rice samples from 1.43 to 3.13%. This result is in agreement with

**Table 1** Nutritional important starch fractions of rice varieties (% dry matter)

Rice varieties	RS	RDS	SDS
<i>FARO 44</i>	2.02 ± 0.03 <sup>bc</sup>	31.57 ± 1.15 <sup>d</sup>	38.58 ± 1.89 <sup>bc</sup>
<i>FARO 52</i>	2.02 ± 0.14 <sup>bc</sup>	35.29 ± 1.61 <sup>abcd</sup>	44.03 ± 3.28 <sup>a</sup>
<i>FARO 57</i>	1.83 ± 0.29 <sup>bc</sup>	37.45 ± 2.38 <sup>ab</sup>	37.28 ± 1.16 <sup>bc</sup>
<i>FARO 60</i>	2.09 ± 0.31 <sup>bc</sup>	36.76 ± 1.98 <sup>ab</sup>	36.09 ± 4.27 <sup>bc</sup>
<i>FARO 61</i>	1.43 ± 0.24 <sup>c</sup>	39.26 ± 3.82 <sup>a</sup>	40.68 ± 0.94 <sup>ab</sup>
<i>Jamila</i>	2.64 ± 0.71 <sup>ba</sup>	27.70 ± 4.64 <sup>e</sup>	35.04 ± 1.28 <sup>c</sup>
<i>Jeep</i>	3.13 ± 1.05 <sup>a</sup>	36.07 ± 1.58 <sup>abc</sup>	39.67 ± 4.52 <sup>abc</sup>
<i>Kwandala</i>	2.74 ± 0.44 <sup>ba</sup>	32.66 ± 1.25 <sup>cd</sup>	40.74 ± 5.12 <sup>ab</sup>
<i>Yardass</i>	2.33 ± 1.00 <sup>abc</sup>	34.83 ± 1.65 <sup>bcd</sup>	38.93 ± 1.90 <sup>bc</sup>

RS, resistant starch; RDS, rapidly digestible starch; SDS, slowly digestible starch.

Mean ± SD values with same superscripts in column are not significantly different ( $P > 0.05$ ).

previous published data on milled nonparboiled rice samples (Hu *et al.*, 2004). Hu *et al.* (2004) reported a range of 0.1–3.2% for RS in nonparboiled, milled rice varieties in China. Goni *et al.* (1997) reported 2.53% RS in rice although that study did not specify whether the samples were milled, nonparboiled rice or milled parboiled. The consistent results in this study with the previous studies can be explained by similar analytical procedures. Walter *et al.* (2005) reported varied mean values of RS (4.2 ± 0.2 and 1.5 ± 0.9) in nonparboiled milled rice samples subjected to different analytical methods. The analytical procedures of Walter *et al.* (2005) differed in terms of sample preparation, pH of buffer, type and concentration of enzyme. For example, methodology that involves protein removal with pepsin treatment enhances amylase accessibility and prevents starch encapsulations by protein matrix (Goni *et al.*, 1996). The RS values of samples treated with pepsin were found to be significantly different from samples without pepsin treatment (Goni *et al.*, 1996).

With respect to the nine rice varieties in our study, highest resistant starch was observed in *Jeep* (3.13%) while lowest value was found in *FARO 61* (1.43%). RDS fraction varied between 27.70% (*Jamila*) and 39.26% (*FARO 61*). The RDS value for *Jamila* was significantly different from all the other rice varieties ( $P < 0.05$ ). With respect to the new rice varieties, *FARO 44* had the lowest RDS value (31.57%), but this was not significantly different from that of *FARO 52* (35.29%). RDS values for *FARO 57*, *FARO 60* and *FARO 61* were also not significantly different from each other ( $P < 0.05$ ). This result suggests that *Jamila* and *FARO 44* have less rapid digestion and glycaemic responses. The SDS fraction ranged between 35.04% (*Jamila*) and 44.03% (*FARO 52*). Levels of SDS and RDS in this study are in agreement with previous



**Table 2** Resistant and digestible starch proportions of rice varieties in raw and cooked samples (% dry matter)

Rice varieties	RS (% of TS)	DS (% of TS)	TS
Cooked samples			
<i>FARO 44</i>	2.02(2.12)	93.11(97.88)	95.13
<i>FARO 52</i>	2.02(2.04)	97.02(97.95)	99.05
<i>FARO 57</i>	1.83(2.11)	84.97(97.88)	86.81
<i>FARO 60</i>	2.09(2.10)	97.48(97.89)	99.58
<i>FARO 61</i>	1.43(1.48)	94.94(98.52)	96.37
<i>Jamila</i>	2.64(2.69)	95.34(97.31)	97.98
<i>Jeep</i>	3.13(3.18)	95.32(96.81)	98.46
<i>Kwandala</i>	2.74(2.93)	90.79(97.07)	93.53
<i>Yardass</i>	2.33(2.71)	83.58(97.29)	85.91
Mean $\pm$ SD	2.29 $\pm$ 0.71 <sup>b</sup>	92.51 $\pm$ 5.56 <sup>a</sup>	94.76 $\pm$ 5.56 <sup>a</sup>
Raw samples			
<i>FARO 44</i>	17.56(30.05)	40.86(69.93)	58.43
<i>FARO 52</i>	16.68(33.38)	33.29(66.62)	49.97
<i>FARO 57</i>	14.92(28.80)	36.89(71.20)	51.81
<i>FARO 60</i>	16.29(32.03)	34.57(67.97)	50.86
<i>FARO 61</i>	13.59(28.15)	34.69(71.85)	48.28
<i>Jamila</i>	16.33(33.75)	32.05(66.23)	48.39
<i>Jeep</i>	13.23(30.45)	30.22(69.55)	43.45
<i>Kwandala</i>	15.01(31.35)	32.87(68.65)	47.88
<i>Yardass</i>	15.74(31.21)	34.69(68.79)	50.43
Mean $\pm$ SD	15.49 $\pm$ 1.85 <sup>a</sup>	34.46 $\pm$ 4.92 <sup>b</sup>	49.94 $\pm$ 5.74 <sup>b</sup>

RS, resistant starch; DS, digestible starch; TS, total starch.

Mean  $\pm$  SD values with same superscripts in column are not significantly different ( $P > 0.05$ ).

reported values of SDS fraction (41.46%) and RDS fraction (34.10%) in milled rice samples (Rosin *et al.*, 2002). High proportion of SDS in food causes a slow increase in postprandial blood glucose levels unlike the RDS that has a fast and high peak of blood glucose response. Therefore, SDS has moderate impact on GI, which is associated with health benefits such as reduced risk of diabetes and cardiovascular disease (Englyst *et al.*, 2003).

There were significant differences ( $P < 0.05$ ) in the total starch, digestible and resistant starch fractions between freshly cooked rice samples and raw samples (Table 2).

The significantly higher TS in cooked samples (85.91–99.58%) compared with raw samples (43.45–51.81%) can probably be explained by gelatinisation of starch during sample preparation (Allen *et al.*, 2012).

The raw rice samples have higher percentage of TS as RS fraction (28.15–33.75%) compared with the freshly cooked samples (1.48–3.18%). This finding of higher RS in raw samples is consistent with previous reports (Sagum & Arcot, 2000; Han *et al.*, 2008). Han *et al.* (2008) reported greater RS content in uncooked flour of rice than in cooked rice grains. Sagum & Arcot (2000) reported a significant reduction of RS

content in three rice varieties after boiling (1.0–2.2%) compared with raw samples (8.6–12.9%) on dry basis. RS fraction can significantly be reduced during cooking by disrupting the crystalline starch structure (Sajilata *et al.*, 2006; Roopa & Premavalli, 2008).

However, the values of RS fraction of raw samples in our present study are higher than RS values reported in a previous study on raw rice (Sagum & Arcot, 2000). The difference in RS values could be attributed to the analytical method adopted in our study which enhanced enzymatic accessibility as samples were subjected to pepsin treatment for protein removal before starch hydrolysis.

The TS of freshly cooked samples were predominantly DS fraction (96.81–98.52%), which is not surprising because the samples were not parboiled before milling. According to Walter *et al.* (2005), DS fractions were higher in nonparboiled rice and lower after parboiling. Similar observation of high digestible starch fraction was reported (97% of TS) in cooked polish rice (Rosin *et al.*, 2002).

The RS of the freshly cooked samples showed a mean value of 2.29%, which was similar to RS (2.41%) observed in the study of Rosin *et al.* (2002). However, the DS (92.51%) and TS (94.76%) in their study were higher than mean values in this study (DS = 84.95%; TS = 87.36%). This difference can be explained by variation in analytical methods for measuring DS and TS. This study used direct method for DS and RS measurements, whereas Rosin *et al.* (2002) determined DS by indirect method. According to Walter *et al.* (2005), indirect methods accumulate more errors from two determinations. It is noteworthy that RDS and SDS fractions in these two studies adopted the same analytical procedure, and thus, values were comparable.

The comparison of starch fractions among varieties (Table 3) showed that local rice varieties had a significantly higher fraction of RS (2.71%) and a lower RDS (32.82%) than the improved varieties (1.88% and 36.07%, respectively). This finding could be attributed to variation in nutritional composition. The relationship between amylose and starch digestibility in rice varieties had been documented (Frei *et al.*, 2003; Hu *et al.*, 2004; Sajilata *et al.*, 2006). In the study by Frei *et al.* (2003), the highest and lowest glycaemic index were found in waxy rice (0% amylose) and high amylose rice (26.9% amylose), respectively. Their study also showed that cultivar with highest RS (1.3%) had a low-glycaemic index. Therefore, a follow-up study on the nutritional value and amylose contents of these rice varieties in our study is needed.

The SDS and the SDI among the local and new rice varieties were not significantly different ( $P > 0.05$ ). Higher RS and lower RDS contents in food are associated with decreased glycaemic responses and

**Table 3** Comparison of starch fractions and starch digestion index (SDI) between local rice varieties and new rice varieties

Starch Parameters	Rice samples		T-value (P-value)
	Improved varieties (%)	Local varieties (%)	
RS	1.88 ± 0.32 <sup>b</sup>	2.71 ± 0.81 <sup>a</sup>	3.882 (P < 0.05)
RDS	36.07 ± 3.39 <sup>a</sup>	32.82 ± 4.07 <sup>b</sup>	2.564 (P < 0.05)
SDS	39.33 ± 3.71 <sup>a</sup>	38.60 ± 3.92 <sup>a</sup>	0.573 (P > 0.05)
SDI	37.94 ± 4.38 <sup>a</sup>	35.14 ± 5.49 <sup>a</sup>	1.660 (P > 0.05)

RS, resistant starch; RDS, rapidly digestible starch; SDS, slowly digestible starch; SDI, starch digestion index.

Mean ± SD values with same superscripts in row are not significantly different (P > 0.05).

**Table 4** Pearson's correlations between study parameters

Starch fractions	RS	SDI
RDS	-0.447**	0.879**
SDS	-0.151NS	0.197NS
RS	1	-0.417*

RS, resistant starch; RDS, rapidly digestible starch; SDS, slowly digestible starch; SDI, starch digestion index; NS, correlation is not significant;

\*correlation is significant at 5% level.

\*\*correlation is significant at 1% level.

significant health benefits in dietary management of people with diabetes mellitus and other metabolic conditions (Jenkins *et al.*, 2002; Brand-Miller *et al.*, 2003; Higgins, 2004).

Table 4 shows an inverse relationship between RS and RDS ( $r = -0.447$ ,  $P < 0.01$ ). RS also had a significantly negative correlation with SDI ( $r = -0.417$ ,  $P < 0.05$ ). This observation is typical as RS fraction is not absorbed in the small intestine and therefore cannot increase the postprandial glycaemic response.

The relative rate of starch digestion as measured by SDI showed a positive and strong correlation with RDS ( $r = 0.879$ ,  $P < 0.01$ ). Similar correlation has been reported in starch digestibility studies (Rosin *et al.*, 2002; Odenigbo *et al.*, 2012).

This highly positive relationship between SDI and RDS is not surprising because the samples in this study predominantly consist of DS that elicits higher glycaemic responses (Larsen *et al.*, 2000; Rosin *et al.*, 2002; Walter *et al.*, 2005). RDS is characterised by rapid digestion and absorption in the small intestine, which is responsible for elevated glycaemic response (Englyst *et al.*, 1999).

## Conclusion

This study reports varied levels of the three nutritionally important starch fractions in milled, nonparboiled

rice. The cooked rice samples were predominantly DS fractions unlike the uncooked samples. Higher levels of RS and lower RDS fraction were found in local varieties compared with the improved varieties. The RS was inversely related to RDS and SDI with a strong positive correlation between RDS and the relative rate of starch digestion (SDI). This information will be useful in the identification of milled, nonparboiled rice varieties with less postprandial glycaemic response.

## Acknowledgments

This study was supported by the Canadian International development Agency (CIDA) project on enhancing food security in Africa through the improvement of rice postharvest handling, marketing and the development of new rice-based products.

## References

- Allen, J.C., Corbitt, A.D., Maloney, K.P., Butt, M.S. & Truong, V.D. (2012). Glycemic Index of Sweet Potato as Affected by Cooking Methods. *Open Nutrition Journal*, **6**, 1–11.
- Brand-Miller, J., Hayne, S., Petocz, P. & Colagiuri, S. (2003). Low-Glycemic Index Diets in the Management of Diabetes. *Diabetes Care*, **26**(8), 2261–2267.
- Dona, A.C., Pages, G., Gilbert, R.G. & Kuchel, P.W. (2010). Digestion of starch: *in vivo* and *in vitro* kinetic models used to characterise oligosaccharide or glucose release – Review. *Carbohydrate Polymers*, **80**, 599–617.
- Eggum, B., Juliano, B., Perez, C. & Acedo, E. (1993). The resistant starch, undigestible energy and undigestible protein contents of raw and cooked milled rice. *Journal of Cereal Science*, **18**(2), 159–170.
- Englyst, H.N. & Hudson, G.J. (1996). The classification and measurement of dietary carbohydrates. *Food chemistry*, **57**(1), 15–21.
- Englyst, H.N., Kingman, S.M. & Cummings, J.H. (1992). Classification and measurement of nutritionally important starch fractions. *European Journal of Clinical Nutrition*, **46**(2), S33–S50.
- Englyst, K.N., Englyst, H.N., Hudson, G.J., Cole, T.J. & Cummings, J.H. (1999). Rapidly available glucose in foods: an *in vitro* measurement that reflects the glycaemic response. *American Journal of Clinical Nutrition*, **69**, 448–454.
- Englyst, K.N., Vinoy, S., Englyst, H.N. & Lang, V. (2003). Glycaemic index of cereal products explained by their content of rapidly and slowly available glucose. *British Journal of Nutrition*, **89**(3), 329–340.
- Frei, M., Siddhuraju, P. & Becker, K. (2003). Studies on the *in vitro* starch digestibility and the glycemic index of six different indigenous rice cultivars from the Philippines. *Food Chemistry*, **83**(3), 395–402.
- Goni, I., Garcia-Diz, L., Mañas, E. & Saura-Calixto, F. (1996). Analysis of resistant starch: a method for foods and food products. *Food Chemistry*, **56**(4), 445–449.
- Goni, I., Garcia-Alonso, A. & Saura-Calixto, F. (1997). A starch hydrolysis procedure to estimate glycemic index. *Nutrition Research*, **17**(3), 427–437.
- Han, J., Jang, S. & Lim, S. (2008). *In vitro* digestibility of rice and barley in forms of raw flour and cooked kernels. *Food Science and Biotechnology*, **17**(1), 180–183.
- Haralampu, S. (2000). Resistant starch—a review of the physical properties and biological impact of RS3. *Carbohydrate polymers*, **41**(3), 285–292.



- Higgins, J.A. (2004). Resistant starch: metabolic effects and potential health benefits. *Journal of AOAC International*, **87**(3), 761–768.
- Hu, P., Zhao, H., Duan, Z., Linlin, Z. & Wu, D. (2004). Starch digestibility and the estimated glycemic score of different types of rice differing in amylose contents. *Journal of Cereal Science*, **40**(3), 231–237.
- Jenkins, D.J.A., Kendall, C.W.C., Augustin, L.S.A. et al. (2002). Glycemic index: overview of implications in health and disease. *American Journal of Clinical Nutrition*, **76**(1), 266S–273S.
- Khatoun, N. & Prakash, J. (2006). Nutritional quality of microwave and pressure cooked rice (*Oryza sativa*) varieties. *Food Science and Technology International*, **12**(4), 297–305.
- Larsen, H.N., Rasmussen, O.W., Rasmussen, P.H. et al. (2000). Glycemic index of parboiled rice depends on the severity of processing: study in type 2 diabetic subjects. *European Journal of Clinical Nutrition*, **54**, 380–385.
- Marsono, Y. & Topping, D.L. (1993). Complex carbohydrates in Australian rice products—Influence of microwave cooking and food processing. *LWT-Food Science and Technology*, **26**, 364–370.
- Newton, J., Wang, Y.J. & Mauromoustakos, A. (2011). Effects of Cultivar and Processing Condition on Physicochemical Properties and Starch Fractions in Parboiled Rice. *Cereal Chemistry*, **88**(4), 414–420.
- Odenigbo, A., Rahimi, J., Ngadi, M., Amer, S. & Mustafa, A. (2012). Starch digestibility and predicted glycemic index of fried sweet potato cultivars. *Functional Foods in Health and Disease*, **2** (7), 280–289.
- Rashmi, S. & Urooj, A. (2003). Effect of processing on nutritionally important starch fractions in rice varieties. *International Journal of Food Sciences and Nutrition*, **54**(1), 27–36.
- Roopa, S. & Premavalli, K.S. (2008). Effect of processing on starch fractions in different varieties of finger millet. *Food Chemistry*, **106**, 875–882.
- Rosin, P.M., Lajolo, F.M. & Menezes, E.W. (2002). Measurement and characterization of dietary starches. *Journal of Food Composition and Analysis*, **15**(4), 367–377.
- Sagum, R. & Arcot, J. (2000). Effect of domestic processing methods on the starch, non-starch polysaccharides and *in vitro* starch and protein digestibility of three varieties of rice with varying levels of amylose. *Food Chemistry*, **70**(1), 107–111.
- Sajilata, M.G., Singhal, R.S. & Kulkarni, P.R. (2006). Resistant starch – A review. *Comprehensive Reviews in Food Science and Food Safety*, **5**, 1–17.
- Tetens, I., Biswas, S.K., Glito, L.V., Kabir, K.A., Thilsted, S.H. & Choudhury, N.H. (1997). Physicochemical characteristics as indicators of starch availability from milled rice. *Journal of Cereal Sciences*, **26**, 355–361.
- Walter, M., Da Silva, L.P. & Denardin, C.C. (2005). Rice and resistant starch: different content depending on chosen methodology. *Journal of Food Composition and Analysis*, **18**(4), 279–285.
- Woolnough, J.W., Monro, J.A., Brennan, C.S. & Bird, A.R. (2008). Simulating human carbohydrate digestion *in vitro*: a review of methods and the need for standardization. *International Journal of Food Science & Technology*, **43**, 2245–2256.

# Mapping of Fat and Moisture Distribution in Atlantic Salmon Using Near-Infrared Hyperspectral Imaging

Fengle Zhu · Hailiang Zhang · Yongni Shao · Yong He · Michael Ngadi

Received: 31 May 2013 / Accepted: 13 November 2013  
© Springer Science+Business Media New York 2013

**Abstract** A nondestructive and rapid method using near-infrared (NIR) hyperspectral imaging was investigated to determine the spatial distribution of fat and moisture in Atlantic salmon fillets. Altogether, 100 samples were studied, cutting out from different parts of five whole fillets. For each sample, the hyperspectral image was collected with a pushbroom NIR (899–1,694 nm) hyperspectral imaging system followed by chemical analysis to measure its reference fat and moisture contents. Mean spectrum were extracted from the region of interest inside each hyperspectral image. The quantitative relationships between spectral data and the reference chemical values were successfully developed based on partial least squares (PLS) regression with correlation coefficient of prediction of 0.93 and 0.94, and root mean square error of prediction of 1.24 and 1.06 for fat and moisture, respectively. Then the PLS models were applied pixel-wise to the hyperspectral images of the prediction samples to produce chemical images for visualizing fat and moisture distribution. The results were promising and demonstrated the potential of this technique to predict constituent distribution in salmon fillets.

**Keywords** Hyperspectral imaging · Salmon · Fat · Moisture · Partial least squares · Chemical image

## Introduction

Salmon has always been regarded as a popular gourmet fish consumed in large quantities due to its high nutritional value. Global aquaculture production for Atlantic salmon (*Salmo salar*) has more than doubled from 805,616 tons in 1999 to 1,721,254 tons in 2011 with a value of US\$9,710,782,000 (FAO 2013). Fat and moisture content in salmon products is closely related to their quality and safety. Concerning eating quality, fresh salmon low or high in fat is often characterized as “dry” and “poor” or “fuller” and “rich” in flavor. Fat content also affect the further processing of salmon flesh, for example, fat content should ideally be between 8 and 12 % for optimum uptake of smoke (Downey 1996). Besides, the shelf life of salmon products is greatly influenced by its moisture content, which affects the microflora growth (Shimasaki et al. 1994). Thereby, the determination of fat and moisture in salmon is significant for both consumers and producers. Conventional wet chemical analysis for fat and moisture are reliable, but they are destructive, lengthy, expensive, and hazardous (Rasco et al. 1991).

During the past few decades, near-infrared (NIR) spectroscopy has been used extensively as a rapid and nondestructive method for compositional determination in fish and fish products (Wold et al. 1996; Khodabux et al. 2007; Folkestad et al. 2008). The absorption between 780 and 2,500 nm arise from combination and overtones bands of C–H, O–H, N–H, and S–H stretching and bending vibrations (Cen and He 2007). Because spectroscopic scan is point measurement, some researchers (Wold et al. 1996; Wold and Isaksson 1997; Folkestad et al. 2008) conducted measurements on Norwegian Quality Cut, corresponding to the flesh from the

---

F. Zhu · H. Zhang · Y. Shao · Y. He (✉)  
College of Biosystems Engineering and Food Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou, Zhejiang 310058, People's Republic of China  
e-mail: yhe@zju.edu.cn

Y. He  
Key Laboratory of Equipment and Informatization in Environment Controlled Agriculture, Ministry of Agriculture, 866 Yuhangtang Road, Hangzhou, Zhejiang 310058, China

M. Ngadi (✉)  
Department of Bioresource Engineering, Macdonald Campus, McGill University, Ste-Anne-de-Bellevue, Québec H9X 3 V9, Canada  
e-mail: michael.ngadi@mcgill.ca

dorsal fin to gut. Chemical contents in this area are good estimates of the mean composition of whole salmon fillet. Nevertheless, according to Gy (2004), this nonprobabilistic sampling method can never generate a truly representative sample. Moreover, fat and moisture are distributed heterogeneously in whole fillet. Generally, the fat content increases from back to belly, from tail to head, and from inside to skin side (Rye 1991). Moisture content negatively correlates with fat (Wold et al. 1996; Wold and Isaksson 1997). Unfortunately, NIR spectroscopy is unable to provide information on spatial distribution of chemical composition.

Hyperspectral imaging integrates spectroscopy and digital imaging to acquire a grayscale image for each band and a spectrum for each pixel (Wu and Sun 2013). By combining the chemical selectivity of spectroscopy with the power of image visualization, it enables a complete description of component concentration and distribution in heterogeneous samples (Gowen et al. 2008). NIR hyperspectral or multispectral imaging has been applied in the determination of fat and moisture in different species of fish fillets (ElMasry and Wold 2008) and fat, salt (Segtnan et al. 2009a, b), and moisture (He et al. 2013) in salmon fillets. All of these publications showed how fat and moisture were distributed in fillets.

In previous studies concerning fat analysis (ElMasry and Wold 2008; Segtnan et al. 2009a, b), after acquiring the hyperspectral image of whole fillet, five cylindrical subsamples (each 10 or 15 mm in diameter) from different locations of fillets were cut out to measure fat content using nuclear magnetic resonance (NMR). Then, spectra from the image pixels corresponding to the circular area where the subsample had been cut out were extracted and averaged to model against its fat reference value. Although good results were achieved in these studies, there were three imperfections in all of them: (1) because of the rather small size of subsamples, there was obvious possible spatial mismatch between excised subsample and delimited circular area in the image for local spectra extraction, which could introduce inaccuracies in the model; (2) the reference fat values were measured by NMR, which is not as classical and accurate as the traditional wet chemical method; (3) only 15 wavelengths were collected between 760 and 1,040 nm, the few number of bands and short-wave NIR spectral region could not provide sufficient information on fat absorption.

In this study, the above imperfections were corrected by the following practices: (1) the whole fillet was first cut into many small pieces, each piece was regarded as a sample. Then, hyperspectral image for every sample was obtained followed by fat and moisture measurement for the corresponding sample. This sampling procedure guaranteed the exact match between the spectral collection and chemical determination for each sample; (2) Soxhlet extraction, a well-known classical method which provides precise and stable results, was employed to determine the reference fat values; (3) the

spectral range used (899–1,694 nm) with 167 bands was more suitable for the detection of organic molecules in fat, resulting in better results for modeling spectral data and fat distribution. The main objective of the present work was to establish a NIR hyperspectral imaging system to determine fat and moisture distribution in salmon fillets, focusing on building robust calibration models to correlate spectral information and chemical values quantitatively, and applying models to all pixels from hyperspectral images in the prediction set to produce distribution maps of fat and moisture along the whole fillet.

## Materials and Methods

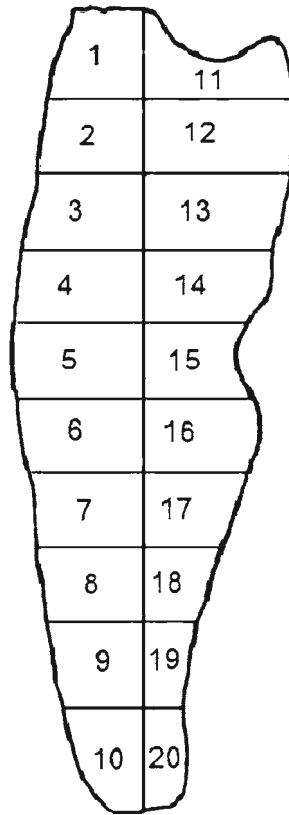
### Fish Samples

Five fresh and superior Atlantic salmon fillets were purchased from the local supermarket (Montreal, Canada). In order to maximize information describing the natural variation in salmon composition, individual fillet were purchased at different times over a period of several months (on Nov 16, Dec 2 and 15, 2012; Jan 19 and Feb 2, 2013, respectively). Fillet weight varied from 920 to 1,188 g (average, 1,079 g) and length ranged between 47 and 52 cm (average, 50 cm). For each whole fillet, dorsal and ventral parts were cut relative to the center line from head to tail; dorsal pieces were numbered 1 to 10 and ventral pieces from 11 to 20 (Fig. 1). Skin was removed in this process. The width of each sample was approximately 5 cm, while the length and thickness varied according to the position where it was taken. Finally, a total of 100 samples were obtained from different locations of five fillets. After cutting, samples were immediately put in individual plastic bags and placed in an insulation box with several dry ice bags around to maintain the constant temperature of 2–4 °C.

### Hyperspectral Image Acquisition and Calibration

The hyperspectral images of samples were captured by NIR hyperspectral imaging system in reflectance mode at 20 °C. The system is composed of a line-scanning spectrograph (Hyperspec™, Headwall Photonics Inc, USA), an InGaAs camera, an illumination system with two 50-W tungsten halogen lamps adjusted at angle of 45° to illuminate the camera's field of view, a conveyer belt (Donner 2200 series, Donner Mfg. Corp, USA) driven by a stepping motor (MDIP22314, Intelligent Motion System Inc, USA) with a user-defined speed, a supporting frame, a computer, and a data acquisition and preprocessing software (Hyperspec, Headwall Photonics Inc, USA) with eight frames to average the output signal and integration time of 2 ms. The speed was set as 1.58 mm/s, selected by trial-and-error to avoid image distortion. The

**Fig. 1** Sampling position across the whole salmon fillet



wavelength range is between 899 and 1,694 nm with spectral resolution of 2.8 nm. The camera has  $167 \times 320$  (spectral  $\times$  spatial) pixels with fixed distance of 40 cm from lens to samples' surface. To acquire a three-dimensional hypercube, each sample was placed on a black background which had very low reflectance and conveyed to the field of view of camera to be scanned line by line. The inner side of wet salmon flesh was facing the camera.

To calibrate the raw hyperspectral image ( $I_0$ ), additional images for white ( $W$ ) and dark ( $B$ ) references were used. The white image was acquired from a standard Teflon white tile ( $\sim 100\%$  reflectance) under the same condition used for samples. The dark image ( $\sim 0\%$  reflectance) was recorded by turning off the light source and completely covering the camera lens with its opaque cap to measure the dark current of camera. The calibrated image ( $I$ ) was calculated by the following equation:

$$I = \frac{I_0 - B}{W - B} \times 100 \quad (1)$$

#### Chemical Analysis

After the collection of hyperspectral images, samples were immediately weighted and kept in the deep freezer ( $-80^\circ\text{C}$ )

overnight. Then, moisture contents were determined gravimetrically by vacuum freeze drying ( $-50^\circ\text{C}$  and 150 mbar) for 48 h. Subsequently, dry samples were ground, and fat contents in the dry matter were measured by Soxhlet extraction method using petroleum ether (AOAC 2006). Each analysis was performed in duplicate.

#### Spectra Extraction and Data Calibration

A polygon region of interest (ROI) which is as large as possible inside each calibrated hyperspectral image was identified in Environment for Visualizing Images (ENVI) software (ITT Visual Information Solutions, USA). The reflectance spectra of all pixels within the ROI were extracted and averaged along each band to generate one mean spectrum for each sample. The same procedure was repeated for all ROI images. Among the total 167 bands, the starting 10 bands and ending 6 bands were removed because of low signal-to-noise ratio, and spectral region between band 11 and 161, corresponding to the wavelength range of 947 to 1,666 nm, was used. This resulted in a spectral matrix of 100 samples  $\times$  151 bands. Four whole fillets with 80 samples were randomly selected to create the calibration set, the remaining 20 samples formed the prediction set.

Partial least squares (PLS) regression was applied to construct the calibration models, with the aid of the software Unscrambler V9.7 (CAMO Process As, Norway). PLS is a widely used bilinear modeling and multi-analysis method (Geladi and Kowalski 1986), which can be considered as a standard calibration technique for spectroscopic analysis. It is especially suitable when the number of variables is greater than the number of samples, and when there is colinearity among variables. The quantitative relationship is built using PLS between independent  $X$  and dependent  $Y$  variables:

$$Y_{n \times 1} = X_{n \times k} \times B_{k \times 1} + E_{n \times 1} \quad (2)$$

where  $X$  is the extracted mean spectral matrix with 151 bands for 80 calibration samples,  $Y$  matrix represents the chemical content of fat or moisture for calibration samples,  $B$  is the matrix of regression coefficients, and  $E$  is the regression residual matrix. Separate models were generated for fat and moisture. Leave-one-out cross-validation was used to validate the quality and evaluate overfitting of the calibration models. Then, the fat or moisture values for samples from prediction set were predicted by the established models. The performances of models were evaluated by correlation coefficient of calibration ( $R_c$ ), cross-validation ( $R_{cv}$ ), prediction ( $R_p$ ), root mean square error of calibration (RMSEC), cross-validation (RMSECV), prediction (RMSEP), and paired-samples  $t$  test between the predicted and reference values for prediction samples. Generally, a good model should have high

**Table 1** Summary of chemical analyses for fat (based on wet weight) and moisture in calibration and prediction samples

Component	Calibration				Prediction			
	Min	Max	Mean	SD	Min	Max	Mean	SD
Fat (%)	3.30	25.13	10.41	4.97	4.44	12.78	8.21	2.49
Moisture (%)	54.89	74.73	67.58	4.26	65.24	72.22	69.13	2.10

R, low RMSE, a small difference among RMSEC, RMSECV, RMSEP, and lower calculated  $t$  than  $t$  critical values.

### Chemical Imaging

Chemical imaging is the technique for creating visual color images to show the quantitative spatial distribution of components in heterogeneous food products. Because of the impossibility to obtain the concrete component concentration for each pixel, a calibration model is first generated with the mean spectra of samples and their corresponding chemical values. The model is then applied to predict component concentration at each pixel in hyperspectral images of the 20 prediction samples, leading to the production of chemical images. Specifically, the spectrum of each pixel within the image was inputted in the established PLS model, i.e., Eq. (2), where  $X$  is the spectral matrix with 151 bands for all pixels. By calculating the dot product between  $X$  and  $B$ , the  $Y$  matrix, representing the fat or moisture content for all pixels, was obtained. The background was segmented with a threshold value of 0.08 in this procedure. Subsequently, different concentrations of the predicted fat or moisture were displayed with specific colors in the resulting chemical image, allowing visualization of the difference in spatial distributions for components on the entire sample. This task was executed with a computer program in Interactive Data Language software (ITT Visual Information Solutions, USA).

## Results and Discussion

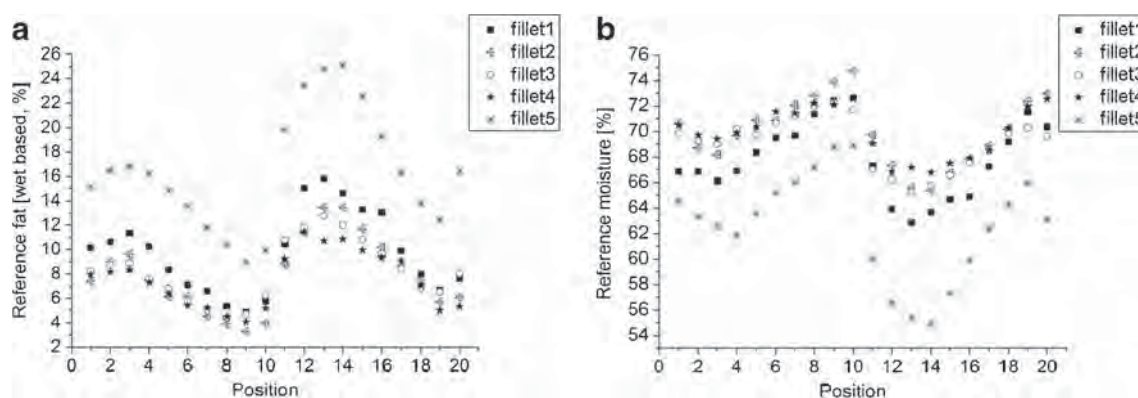
### Reference Chemical Analyses

Table 1 shows a summary of chemical analyses for fat (based on wet weight) and moisture in both calibration and prediction samples. There is a large variation in the reference chemical contents, ranging from 3.30 to 25.13 % for fat and from 54.89 to 74.73 % for moisture. The chemical ranges were similar to that reported in Downey (1996), Wold et al. (1996), Wold and Isaksson (1997), Folkestad et al. (2008), and Segtnan et al. (2009a, b) for fat in salmon and in Downey (1996), Wold and Isaksson (1997), He et al. (2013) for moisture in salmon, indicating that samples with a wide compositional distribution were collected, which was critically important to develop stable calibration models.

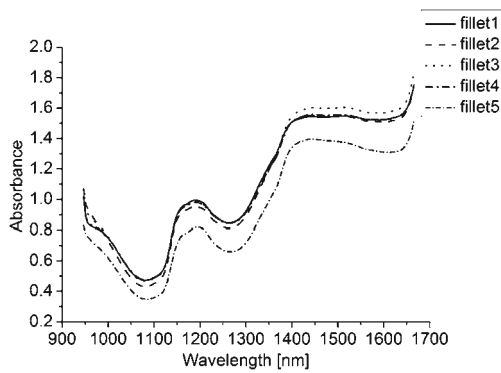
The fat and moisture values as a function of sampling position on fillets are displayed in Fig. 2 (position numbers refer to Fig. 1). The variations of both fat and moisture in different fillets were similar despite of the diverse levels of compositional values. From Fig. 2a, the fat contents were generally higher on the ventral side (positions 11–20) compared with the dorsal side (positions 1–10). And for both dorsal and ventral sides, the fat contents generally decreased from head to tail, except the slight increment of fat contents in the first three positions and the last one position. The variation of moisture contents (Fig. 2b) in both dorsal and ventral sides was just the opposite to that of fat contents, due to the natural high negative correlation between fat and moisture in salmon, in this study  $r = -0.97$ .

### Overview of the Spectra

The NIR absorbance ( $\log 1/R$ ) spectra from 20 samples of each whole fillet were averaged (Fig. 3). The general trends of the spectral curves for five fillets were similar. The absorbance of the fifth fillet was lower than others because of its lower

**Fig. 2** Chemical fat (a) and moisture (b) values as a function of sampling position on fillets





**Fig. 3** Averaged NIR absorbance (log 1/R) spectra of 20 samples of each whole fillet

moisture contents, which was the primary component in salmon. The absorption bands around 1,210 and 1,650 nm were attributed to the C–H stretch second and first overtones in fat, respectively. The wavelengths near 960 and 1,450 nm were assigned to moisture absorption, corresponding to the second and first overtones of O–H stretching and bending, respectively. Moreover, the band at approximately 1,080 nm was related to the second overtone N–H stretching (Rasco et al. 1991). The absorptions of fat and moisture in NIR spectral range are the theoretical basis for establishing the mathematical correlation between spectral data and compositional contents.

### PLS Models

The performances of PLS regression models using reflectance spectra for fat and moisture analysis are presented in Table 2 and Fig. 4. Different spectral transformations were also tested in modeling, showing that reflectance, absorbance, Savitzky–Golay first and second derivatives of reflectance gave approximately the same models and prediction results (data not shown).

For fat model, similar satisfactory results were achieved in calibration ( $R_c=0.96$ , RMSEC=1.21 %) and cross-validation ( $R_{cv}=0.95$ , RMSECV=1.40 %). The absolute difference between  $R_c$  and  $R_p$  (0.93) was 0.03, between RMSEC and RMSEP (1.24 %) was 0.03, respectively. Compared with other publications for fat determination, model errors were lower than the RMSEC of 2.70 %, RMSECV of 2.99 % in ElMasry and Wold (2008), and the RMSECV of 1.96 and 1.95 % for raw and salted fillets, respectively, in Segtnan et al.

(2009a, b). Similarly, a good model for moisture was also obtained with high  $R_c$ ,  $R_{cv}$ , and  $R_p$  of 0.97, 0.95, and 0.94, respectively, and small difference among RMSEC (0.86 %), RMSECV (1.10 %), and RMSEP (1.06 %). In comparison with previous works for moisture prediction, model errors were only 35.10 % of RMSEC and 40.29 % of RMSECV in ElMasry and Wold (2008; RMSEC=2.45 %, RMSECV=2.73 %), and only 55.84 % of RMSEC, 66.67 % of RMSECV and 66.67 % of RMSEP reported by He et al. (2013; mean RMSEC=1.54 %, RMSECV=1.65 %, RMSEP=1.59 %). In addition,  $t$  values of 1.82 and 1.41 for fat and moisture were obtained, respectively. The two-sided  $t$  critical value at 5 % significance level for 19 degrees of freedom was 2.09, suggesting no significant difference between predicted and reference values for both models. The above accurate results confirmed the suitability of hyperspectral imaging for fat and moisture prediction in salmon.

### Chemical Images for Fat and Moisture

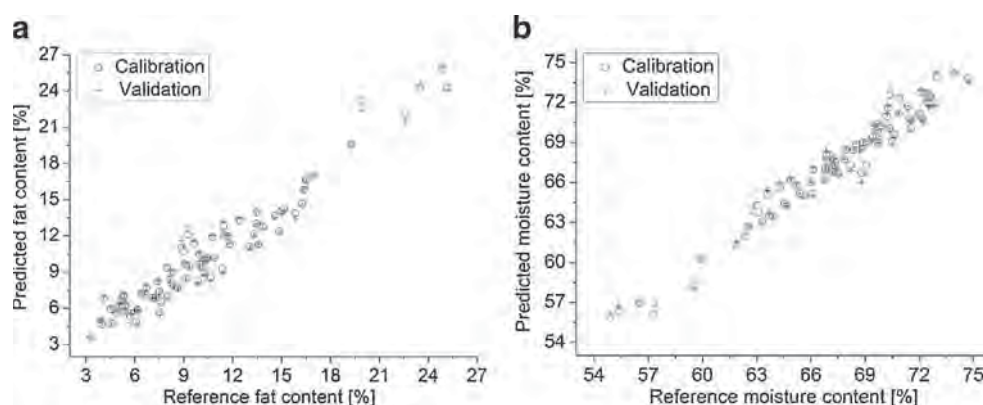
The chemical images of fat and moisture for the 20 prediction samples are shown in Fig. 5 (sample positions refer to Fig. 1). The reference chemical values were marked beside each sample. The changes in fat and moisture contents were assigned with linear color scales. Different colors represented different concentrations of the predicted chemical compositions. Pixels having similar spectral patterns in original hyperspectral images would produce similar predicted values of fat or moisture, and then would appear in similar colors in the resulting chemical images.

As shown in Fig. 5, fat and moisture contents varied greatly among different positions of the same whole fillet and even within samples. The 20 chemical images in Fig. 5a revealed the changing tendency of the predicted fat contents, with fat increasing from dorsal to ventral side, and from tail to head, in agreement with the reference fat variation in Fig. 2a. The changing tendency of the predicted moisture contents in Fig. 5b was contrary to that of the predicted fat contents in Fig. 5a, while the same to that of the reference moisture values in Fig. 2b. Moreover, due to the heterogeneity of compositional contents within one sample, each chemical image was characterized by non-uniform multicolor. Though it was infeasible to measure the chemical values for every point, the high performances of PLS model for composition prediction,

**Table 2** Results of PLS regression models for fat and moisture in salmon fillets

Component	Calibration		Cross-validation		Prediction			
	$R_c$	RMSEC (%)	$R_{cv}$	RMSECV (%)	$R_p$	RMSEP (%)	$t$	$t_{(19, 0.05)}$
Fat	0.96	1.21	0.95	1.40	0.93	1.24	1.82	2.09
Moisture	0.97	0.86	0.95	1.10	0.94	1.06	1.41	2.09

**Fig. 4** Reference and predicted fat (a) and moisture (b) contents in salmon fillets by PLS models



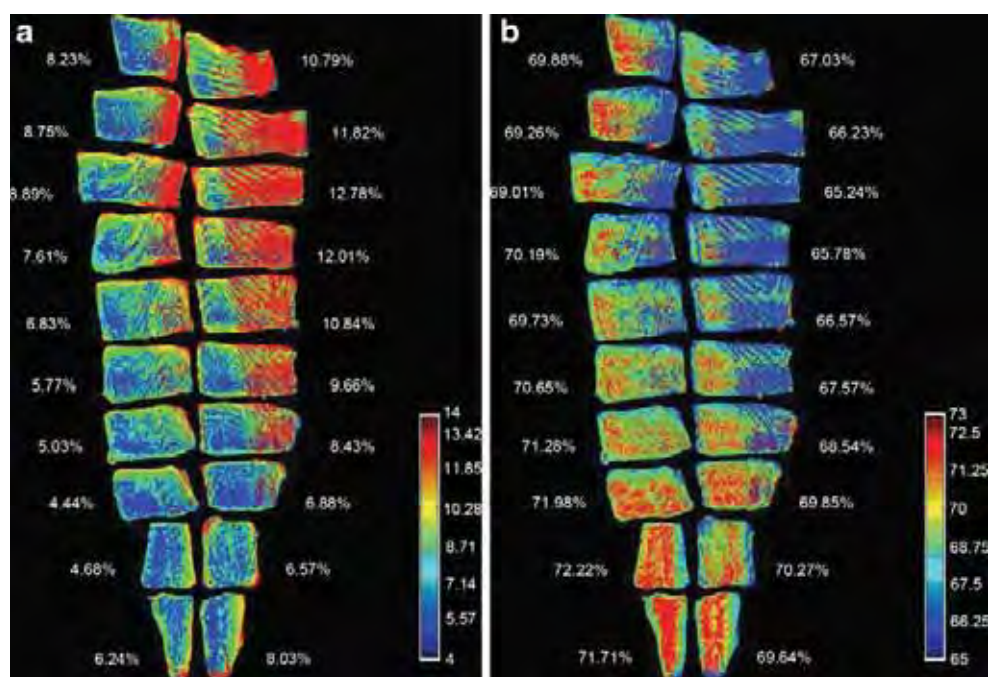
as well as the excellent interpretation on fat and moisture variation along the whole fillet ensured the reliable prediction of compositional contents for every point.

As only one whole fillet was used as prediction samples, fat and moisture ranges of the chemical images were not wide enough to represent the variations in reality. In further work, including larger number of samples with broader chemical values should be carried out for developing more adequate models to produce chemical images for whole fillets, which could enable sorting of whole fillets. For instance, fat images with more red, yellow and green, blue pixels could be classified to fatter, medium, leaner fillets, respectively. Moreover, based on their fat images, whole fillets could be cut into very low (<5%), low (5–9%), medium (9–13%), high (13–17%), and very high (>17%) fat sections, which would facilitate products grading and further processing in fish industry. This optimized trimming cannot be implemented based on point

measurement using conventional spectroscopy due to the heterogeneous compositional distribution on whole fillet. Hence, it is advantageous to apply hyperspectral imaging to display more detailed and accurate information on how and where different concentrations of fat and moisture are distributed. However, besides the necessity of collecting more samples for calibration, some limitations of the proposed technique must be addressed prior to its practical application.

The primary limitation is the required time for image acquisition and processing (40 s for one sample) is too great for online detection, although it is much shorter than chemical analysis. To solve this problem, a few optimal wavelengths for reducing data dimensionality should be selected to construct a multispectral imaging system. Another limitation derives from the spectral detection depth of the NIR optical system, which is approximately 15 mm down into the fillet (Segtnan et al. 2009a). For large fillet with thickness of more than 15 mm,

**Fig. 5** Chemical images of fat (a) and moisture (b) for the 20 samples in prediction set



fish flesh close to skin in the thick lion area may not be detected, leading to the low estimation in fat of loin part because more fat is located toward the skin and less to the inner side. It might therefore be advantageous to include different fillet thickness in models. Furthermore, since spectral measurements are affected by the variation in temperature, and the temperature in industry is more likely to be close to 2–4 °C, which might differ from that in lab conditions, in further study temperature correction should be included in calibration to assure insensitivity of models to temperature variation. Some approaches were reported to achieve this (Chen and Martin 2007). In addition, more attention should be paid on the frequent calibration of the system (every few hours) in future investigations.

## Conclusion

Mapping of fat and moisture distribution in salmon was achieved using NIR hyperspectral imaging. This study not only corrected the three imperfections found in some previous studies, but also achieved sound modeling results and satisfactory chemical images, and pointed out some limitations of the proposed technique which will be conducive to further research. Fish industry can benefit from adopting this technique with real-time documentation of composition in a pixel-wise manner, proper sorting, and cutting of fillets with certain concentration thresholds and optimal labeling and pricing.

**Acknowledgments** This study was supported by 863 National High-Tech Research and Development Plan (2012AA101903), Fundamental Research Funds for the Central Universities, and funding from the Natural Science and Engineering Research Council (NSERC) of Canada.

## References

AOAC (2006). Official method 991.36. Association of Official Analytical Chemists, Washington, D.C.

Cen, H., & He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science & Technology*, *18*, 72–83.

Chen, T., & Martin, E. (2007). The impact of temperature variations on spectroscopic calibration modelling: a comparative study. *Journal of Chemometrics*, *21*, 198–207.

Downey, G. (1996). Non-invasive and non-destructive percutaneous analysis of farmed salmon flesh by near infra-red spectroscopy. *Food Chemistry*, *55*, 305–311.

ElMasry, G., & Wold, J. P. (2008). High-speed assessment of fat and water content distribution in fish fillets using online imaging spectroscopy. *Journal of Agricultural and Food Chemistry*, *56*, 7672–7677.

FAO. (2013). Fisheries and Aquaculture Information and Statistics Service. <http://www.fao.org/fishery/statistics/global-aquaculture-production/en>.

Folkestad, A., Wold, J. P., Rørvik, K. A., Tschudi, J., Haugholt, K. H., Kolstad, K., & Mørkøre, T. (2008). Rapid and non-invasive measurements of fat and pigment concentrations in live and slaughtered Atlantic salmon (*Salmo salar* L.). *Aquaculture*, *280*, 129–135.

Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, *185*, 1–17.

Gowen, A. A., O'Donnell, C. P., Cullen, P. J., & Bell, S. E. J. (2008). Recent applications of chemical imaging to pharmaceutical process monitoring and quality control. *European Journal of Pharmaceutics and Biopharmaceutics*, *69*, 10–22.

Gy, P. (2004). Sampling of discrete materials—a new introduction to the theory of sampling: I. Qualitative approach. *Chemometrics and Intelligent Laboratory Systems*, *74*, 7–24.

He, H.-J., Wu, D., & Sun, D.-W. (2013). Non-destructive and rapid analysis of moisture distribution in farmed Atlantic salmon (*Salmo salar*) fillets using visible and near-infrared hyperspectral imaging. *Innovative Food Science & Emerging Technologies*, *18*, 237–245.

Khodabux, K., L'Omelette, M. S. S., Jhaumeer-Laulloo, S., Ramasami, P., & Rondeau, P. (2007). Chemical and near-infrared determination of moisture, fat and protein in tuna fishes. *Food Chemistry*, *102*, 669–675.

Rasco, B. A., Miller, C. E., & King, T. L. (1991). Utilization of NIR spectroscopy to estimate the proximate composition of trout muscle with minimal sample pretreatment. *Journal of Agricultural and Food Chemistry*, *39*, 67–72.

Rye, M. (1991). Prediction of carcass composition in Atlantic salmon by computerized tomography. *Aquaculture*, *99*, 35–48.

Segtnan, V. H., Høy, M., Lundby, F., Narum, B., & Wold, J. P. (2009a). Fat distribution analysis in salmon fillets using non-contact near infrared interactance imaging: a sampling and calibration strategy. *Journal of Near Infrared Spectroscopy*, *17*, 247–253.

Segtnan, V. H., Høy, M., Sørheim, O., Kohler, A., Lundby, F., Wold, J. P., & Ofstad, R. (2009b). Noncontact salt and fat distributional analysis in salted and smoked salmon fillets using X-ray computed tomography and NIR interactance imaging. *Journal of Agricultural and Food Chemistry*, *57*, 1705–1710.

Shimasaki, T., Miake, K., Tsukamasa, Y., Sugiyama, M., Minegishi, Y., & Shinano, H. (1994). Effect of water activity and storage temperature on the quality and microflora of smoked salmon. *Nippon Suisan Gakkaishi*, *60*, 569–576.

Wold, J. P., & Isaksson, T. (1997). Non-destructive determination of fat and moisture in whole Atlantic salmon by near-infrared diffuse spectroscopy. *Journal of Food Science*, *62*, 734–736.

Wold, J. P., Jakobsen, T., & Krane, L. (1996). Atlantic salmon average fat content estimated by near-infrared transmittance spectroscopy. *Journal of Food Science*, *61*, 74–77.

Wu, D., & Sun, D.-W. (2013). Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: a review—part I: fundamentals. *Innovative Food Science & Emerging Technologies*. doi:10.1016/j.ifset.2013.04.014.



## Optimization of microencapsulation of probiotics in raspberry juice by spray drying

Kartheek Anekella<sup>a</sup>, Valérie Orsat<sup>b,\*</sup>

<sup>a</sup> Department of Food, Bioprocessing and Nutrition Sciences, Schaub Hall, North Carolina State University, Raleigh, NC 27695, USA

<sup>b</sup> Bioresource Engineering Department, McGill University, 21111 Lakeshore Rd., Ste-Anne de Bellevue, QC, Canada H9X 3V9

### ARTICLE INFO

#### Article history:

Received 27 April 2012  
Received in revised form  
31 July 2012  
Accepted 2 August 2012

#### Keywords:

Sub-lethal thermal effect  
Probiotic/prebiotic  
*Lactobacillus*

### ABSTRACT

**Aims:** Probiotics were microencapsulated in raspberry juice through spray drying.

**Methods & Results:** A combination of probiotics (*Lactobacillus acidophilus* NRRL B-4495 and *Lactobacillus rhamnosus* NRRL B-442) was chosen to offer high viability. Maltodextrin's role as a carbon source was also assessed for its prebiotic potential. Spray drying inlet temperature (°C), total solids: maltodextrin ratio, and inlet feed rate (mL/min) were fixed as independent variables while % recovery, % survival and color were the dependent outputs.

**Conclusions & Significance:** High temperatures during spray drying are detrimental to probiotics and can be circumvented by sub-lethal thermal shock (50 °C for *L. acidophilus* and 52.5 °C for *L. rhamnosus*). Increasing the microencapsulating material concentration increased the survival rate of the probiotics. Non-dairy probiotic foods are becoming popular as they do not pose problems of lactose intolerance while they offer an alternative.

© 2012 Elsevier Ltd. All rights reserved.

### 1. Introduction

Probiotics are considered as “good buddies” to human health. Although historically, probiotics were products of the pharmaceutical industry, the current trend is moving toward the health food sector, making Hippocrates' statement “*Let food be your medicine*” true once again.

Probiotics are usually sold as capsules, powders and combinations of different species which may have multiple advantages (Timmerman, Koning, Mulder, Rombouts, & Beynen, 2004). It was proven in various animal studies that the use of one or more strains/species of probiotics can have beneficial effect and it is logical to assume that a mosaic of probiotics could help in exerting multiple benefits they possess (Famularo, de Simone, Matteuzzi, & Pirovano, 1999; Sanders, 1999). In many cases however, the functionality of probiotics is an issue due to the poor quality in the standards of preparation of probiotics foods and lack of sound and thorough clinical evidence (Azcarate-Peril, Tallon, & Klaenhammer, 2009; Hamilton-Miller & Shah, 2002; Klaenhammer, 2000; Timmerman et al., 2004). The main goal of any food industry is to increase the versatility in consumption of different forms of food

without losing its basic properties of providing nutrition and ensuring health.

Formulating and enriching foods with probiotics would not only improve public health but also the diversity in food choices. Most probiotic foods in the current market are refrigerated dairy-based products (Burgain, Gaiani, Linder, & Scher, 2011) while preparations of non-dairy foods will attract a broader range of consumers with different preferences. The food matrix encapsulation must act as a buffer during storage as well as in the stomach transit until the probiotic is delivered to the intestinal tract along with offering a protection during thermal processing (Ranadheera, Baines, & Adams, 2010). The probiotic microorganisms present in food should survive in significant numbers of at least  $10^6$ – $10^8$  CFU/g, although the numbers vary from strain to strain (Boylston, Vinderola, Ghoddusi, & Reinheimer, 2004; Chávez & Ledebøer, 2007; Ishibashi & Shimamura, 1993; Kailasapathy & Rybka, 1997). It is possible to induce a sub-lethal effect on microorganisms which adapts them to adverse conditions during drying, storage and other processes (Broadbent & Lin, 1999). Usually a temperature rise of 10 °C above the optimal growth temperature leads to shock (Teixeira, Castro, & Kirby, 1994). Thermal sub-lethal treatment can increase the survival rate of *Lactobacilli* remarkably (between 16 and 18 folds depending on the adaptation media) during and following spray drying (Desmond, Stanton, Fitzgerald, Collins, & Paul Ross, 2001; Gardiner et al., 2002). The stress resistance proteins are produced

\* Corresponding author.

E-mail addresses: [kanel@ncsu.edu](mailto:kanel@ncsu.edu) (K. Anekella), [valerie.orsat@mcgill.ca](mailto:valerie.orsat@mcgill.ca) (V. Orsat).



mainly during the sub-lethal exposure prior to drying (Dabbah, Moats, & Mattick, 1969; Teixeira et al., 1994).

Although spray drying of *Lactobacillus* cultures was first done in 1914 by Rogers, it was not adopted due to very low survival rate, difficulty in storage as well as poor rehydration capacity (Porubcan & Sellers, 1975; Teixeira, Castro, Malcata, & Kirby, 1995). Use of berries with probiotics has been tested for uropathogenic and urogenital disorders. The proanthocyanins present in berries, such as cranberries, can modulate the immune system in conjunction with probiotics (Reid, 2002). Anthocyanin rich raspberries also have a high amount of dietary fibers (6.5 g/100 g) with good absorption characteristics which could potentially serve as a carrier and microencapsulating agent as well as a prebiotic (Chiou & Langrish, 2007). This supports our premise of studying the combination of probiotics organisms with raspberry juice for spray drying. Additionally, thermoplasticity and hygroscopicity of fruit juice might pose problems during the spray drying causing them to adhere to the chamber wall due to their stickiness, clogging and caking (Chegini & Ghobadian, 2005). Adding suitable wall materials such as maltodextrin can reduce the caking and stickiness to the walls and increase the free flowing nature of the spray dried powder.

## 2. Materials and methods

Following a review of the literature on the health benefits of probiotics, two species widely used in commercial probiotic foods were chosen. *Lactobacillus rhamnosus* NRRL B-4495 and *Lactobacillus acidophilus* NRRL B-442 were obtained from the USDA's Agricultural Research Services Culture Collection. Dried pellets were reconstituted in 50 mL MRS broth (deMan, Rogosa and Sharpe medium, Difco) and grown overnight (14–16 h) in an incubator shaker (Model G24, New Brunswick Scientific, USA) at 110 rpm and 37 °C. This was subcultured and grown overnight again. The culture thus obtained after the second sub-culture was used for further experiments. A small part of the culture was stored in sterile 30mL/100 mL of glycerol at –80 °C (–86C ULT Freezer Model 5698, Thermo Forma, USA) for later usage. A similar recipe was adapted to prepare MD-MRS medium where the dextrose was replaced by maltodextrin (MD, referred as MD-MRS) with Dextrose Equivalent value 5–8 (Oxoid, USA).

### 2.1. Growth curve and dry biomass estimation

The increase in biomass over the growth cycle period provides information on the kinetics of the growth in relation to substrate utilization. A 5 mL sample of this culture was taken every 3 h and centrifuged in a pre-weighed centrifuge at 1130 × g for 6–7 min. The supernatant was decanted and stored for substrate estimation. The weight of dry biomass was measured as the weight difference between the tube and the pellet after drying at 103 ± 2 °C in hot air oven (Thermoelectron, USA) for over 12 h (Eq. (1)).

$$K = [\log(X_{t1}) - \log(X_{t2})] / 0.301 * (t2 - t1) \quad (1)$$

K = number of populations doubling in an hour (growth rate constant)

$X_{t1}$  = number of cells/mL at time t1;  $X_{t2}$  = number of cells/mL at time t2

During any phase, the doubling time (Td) or generation time is the time required for the cell biomass to double in number as represented in Eq. (2).

$$Td = 1/K \quad (2)$$

### 2.2. Determination of substrate concentration (anthrone method)

Understanding the substrate utilization pattern of a microorganism is also helpful to assess the growth behavior and survival in the gut which is the final destination (Verdenelli et al., 2009). A procedure similar to the one described by Sadasivam and Manickam (2005) was followed. Anthrone reagent (Sigma Aldrich, USA) was prepared by dissolving 200 mg of anthrone in 100 mL of 95 g/100 g sulfuric acid. A 1 mL volume of the supernatant obtained from the biomass estimation was diluted 10 times and 4 mL anthrone reagent was added. The mixture was heated in a boiling water bath (100 ± 2 °C) for 6–7 min until a stable green color was obtained after which the mixture was cooled immediately in an ice bath for 2 min. The maximum absorbance at OD<sub>620</sub> was recorded. The concentration of pure dextrose/maltodextrin was estimated from the measured OD<sub>620</sub> using the standard curve obtained using a known concentration (0.1 g/L) of dextrose/maltodextrin. OD<sub>620</sub> was plotted on the X-axis and concentration of substrate on the Y-axis.

### 2.3. Raspberry juice

The extracted raspberry pulp had a solid content of 13–14 °Brix measured by a portable refractometer (Fischer Scientific, USA). The seeds and skin were sieved out using a fine metal sieve filter. The °Brix unit was adjusted to 11 (total solid concentration 0.1 g/L) as otherwise the pure extract was too viscous to be spray dried.

### 2.4. Sub-lethal temperature ( $T_{sl}$ ) treatment

A sub-lethal temperature treatment was selected to subject the microbes to thermal stress before spray drying. Since the sub-lethal temperature of any given microorganism is strain-specific, each strain was tested individually. Five mL of late log phase *Lactobacillus* cultures (incubated at 37 °C, on a rotary shaker at 110 rpm for 12–14 h) was subjected to different sub-lethal temperatures (45 °C, 50 °C, 52 °C, 55 °C) in a hot water bath. Test tubes containing either MRS or raspberry juice (with inserted thermocouples) were used to regulate the temperature. One mL of sample was collected every 5 min and transferred into 9 ml of sterile MRD (maximum recovery diluents; 1 g/L- peptone, 8.5 g/L NaCl) up to 15 min. CFU analysis was performed from each of these samples at regular intervals (0, 5, 10, 15 min).  $T_{sl}$  was determined from the curves in the graph where the temperature was retaining the highest number of cells after 15 min of thermal stress exposure. Prepared graphs represent the average of triplicate Petri plates for each of three trials.

### 2.5. Spray drying

Spray drying of raspberry juice with maltodextrin as an additive (wall material) at different ratios and mixture of lactobacilli was performed using a Buchi B-290 mini spray dryer. A five mL volume of each probiotic cultures (at late exponential phase, grown overnight) were subjected to sub-lethal treatment and centrifuged at 1130 × g for 1–2 min in order to extract the pure cells. These pellets were diluted in 1 mL of sterile distilled water each and mixed with 50 mL of raspberry pulp and maltodextrin thoroughly by magnetic stirring just before spray drying. Total number of cells just before spray drying was approximately 9.5 log CFU/mL. The spray dryer was allowed to reach uniform process temperature for 15–20 min prior to the spray drying. The aspiration was maintained at 100% and cyclone air flow rate at 30 m<sup>3</sup>/h. The temperatures used to optimize the model were 100 °C, 115 °C and 130 °C. Feed rates used



**Table 1**  
Spray drying responses (generated by JMP-SAS) with the outputs for each dependent variables.

Response	Pattern	Inlet temp (°C)	Maltodextrin ratio	Inlet feed rate (mL/min)	Outlet temperature <sup>a</sup> (°C)	% recovery	% survival	Color (ΔE)
R1	0,0,0	115	1.5	50	82–86	25.4	71.14	54.252
R2	0,0,0	115	1.5	50	80–85	32.9	84.44	54.094
R3	0,0,-1	115	1.5	40	81–86	36	68.24	52.766
R4	+1,-1,-1	130	1	40	92–97	55	58.80	56.719
R5	-1,+1,+1	100	2	60	69–74	25.6	78.97	52.187
R6	0,0,0	115	1.5	50	80–85	35.2	69.10	52.149
R7	-1,-1,+1	100	1	60	67–72	28	80.26	56.623
R8	0,+1,0	115	2	50	81–85	31.1	64.38	52.689
R9	0,0,0	115	1.5	50	81–86	35.2	71.46	52.062
R10	-1,+1,-1	100	2	40	71–76	32.1	82.62	55.443
R11	-1,-1,-1	100	1	40	68–74	47.1	84.33	56.394
R12	-1,0,0	100	1.5	50	71–76	30.75	79.08	53.313
R13	+1,0,0	130	1.5	50	91–95	36.8	53.65	53.413
R14	0,0,0	115	1.5	50	83–88	27.6	67.70	54.337
R15	+1,+1,-1	130	2	40	91–96	38	53.65	51.558
R16	0,-1,0	115	1	50	79–83	41.5	69.74	57.082
R17	+1,+1,+1	130	2	60	88–92	32.5	68.03	53.238
R18	+1,-1,+1	130	1	50	90–94	35.5	56.87	58.018
R19	0,0,+1	115	1.5	60	77–83	24.65	70.39	53.165
R20	0,0,0	115	1.5	50	80–85	27.6	67.70	54.337

<sup>a</sup> Outlet Temp is neither of the variables. It was only included in the table to compare how it changed with respect to the other variables.

were 10, 15, 20 mL/min and the total solids: maltodextrin ratios (referred as maltodextrin ratio from here on) used were 1:1, 1:1.5 and 1:2.

Following spray drying, the raspberry powder thus obtained was stored in small glass jars with screw caps to evaluate their end-product color.

## 2.6. Color

Color of the end products was measured by Chromameter (Model- CR300, Konica-Minolta, USA). The measurement was done in small glass tube filled with the spray dried powder and reconstituted liquid, fully covering the base diameter of the instrument. From the obtained  $L^*$ ,  $a^*$  and  $b^*$  values, the total change in color ( $\Delta E$ ) was calculated using Eq. (3).

$$\Delta E = \sqrt{(L_o - L^*)^2 + (a_o - a^*)^2 + (b_o - b^*)^2} \quad (3)$$

Where the measurements made with the subscript o are made with the reference white calibration. Thus the most desirable bright red color of our raspberry spray dried powder would be indicated in terms of the largest  $\Delta E$  color difference.

## 2.7. Response surface design

Process optimization for the spray drying of probiotics in raspberry juice was performed using response surface methodology (RSM). Initial trials were performed by varying the different conditions for the spray drying of probiotics in raspberry juice such as inlet temperature, maltodextrin ratio, and the inlet feed rate as found in the literature (preliminary results not shown). Inlet temperature (°C), maltodextrin ratio, feed rate (mL/min) were the independent variables and % recovery, % survival and color ( $\Delta E$ ) were the dependent variables. After analyzing the effects of different variables on the probiotics viability, powder color and powder recovery, the maximum and minimum values of the each factor were adjusted for heat shock treated probiotic mixture in the raspberry juice. A central composite design- uniform precision model by JMP-8 (SAS Institute, NC, USA) generated the 20 responses shown in Table 1.

## 3. Results

### 3.1. Growth curves

The organisms' growth curves and their characteristics are presented in Fig. 1 and Table 2 for *L. acidophilus* and in Fig. 2 and Table 3 for *L. rhamnosus*. The relation between OD<sub>600</sub> and log CFU/mL for *L. acidophilus* in MRS is given by the empirically derived equation  $y = 0.74x + 7.81$ . The relation between OD<sub>600</sub> and log CFU/mL for *L. rhamnosus* is given by the empirically derived equation  $y = 1.91x + 7.23$ . These relations were used to estimate the concentration of live cells in MD-MRS from OD<sub>600</sub> values obtained under the same growth conditions.

Although the growth was not very strong with maltodextrin (MD-MRS), as with dextrose (MRS), maltodextrin could still potentially act as a prebiotic as well as a microencapsulating agent during spray drying. From the data in Tables and Figs. 1 and 2, it is concluded that dextrose was utilized better as a carbon source than maltodextrin at similar concentrations. Under similar culture conditions for the two growth media, the residual concentration of dextrose was 6.8 g/L for *L. acidophilus* and 5.5 g/L for *L. rhamnosus* (Figs. 3 and 4). The residual maltodextrin was 11.1 g/L for *L. acidophilus* and 13.8 g/L for *L. rhamnosus* (Figs. 3 and 4).

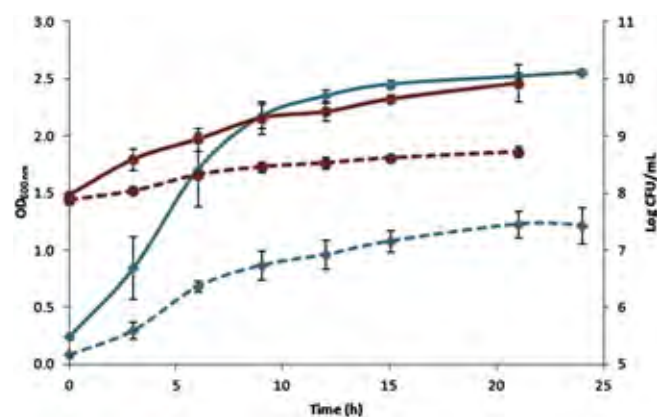


Fig. 1. OD<sub>600</sub> and log CFU of *L. acidophilus* over time in MRS and MD-MRS (37 °C, 110 rpm).

**Table 2**  
Growth curve characteristics of *L. acidophilus* in MRS and MD-MRS.

	MRS	MD-MRS
Growth Curve Features	Exponential phase 0–13th h followed by stationary phase	Exponential phase till 6th h followed by stationary phase
Increase in log CFU/mL	1.75	0.84
$K_{max}$	0.697	0.313
Td	1.43 min	3.2 min

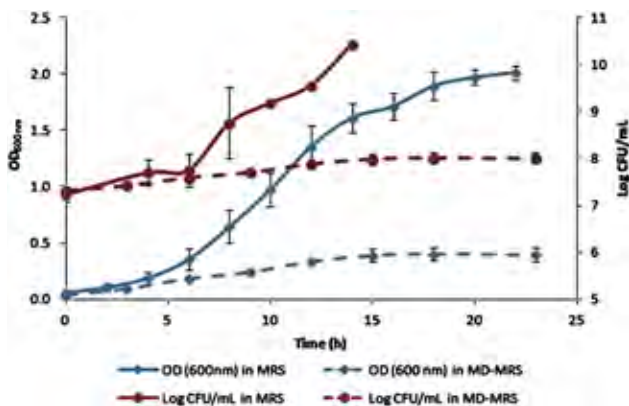
### 3.2. Sub-lethal temperature treatment

Exposing the probiotics to sub-lethal thermal shock increases the subsequent tolerance to near lethal thermal stresses. In the current study, two different heating media were used in the assessment: MRS broth and raspberry juice on the two species separately. A fruit juice was selected, in this case raspberry juice, because it obviates the extra steps of centrifugation and purification of the cultures before spray drying which was done in the case of MRS. Spray drying trials with fruit juice with probiotic cells without heat shock treatment had lower viability than heat shock treated cells under similar conditions.

Figs. 5 and 6 present the organisms' survival during sub-lethal treatment. It can be seen that the viability, in MRS broth, of the probiotics was maintained up to 50 °C for *L. acidophilus* and 52.5 °C for *L. rhamnosus* for 12 min and hence these temperatures have been considered as their sub-lethal temperatures. Raspberry was not as effective as MRS in survival of *L. acidophilus* as the cell count decreased to zero within 5 min at 45 °C and *L. rhamnosus* at 50 °C. The heating medium in which the cells are being adapted has a significant effect on the survival, where a complex media offers a better protection than a simple medium (here raspberry pulp) (Corcoran, Ross, Fitzgerald, & Stanton, 2004; Desmond et al., 2001; O'Riordan, Andrews, Buckle, & Conway, 2001; Teixeira, Castro, & Kirby, 1995).

### 3.3. Spray drying process optimization using RSM

The spray dried powder recovery varied between 25% and 55% depending on the maltodextrin ratio and feed rate. The response plot for % recovery with respect to total solids: maltodextrin ratio and feed flow rate is presented in Fig. 7. The grid on the top indicates the overall optimum condition with respect to maximum % recovery (48.79%) at input conditions of 100 °C inlet temperature,



**Fig. 2.** OD<sub>600nm</sub> and Log CFU of *L. rhamnosus* over time in MRS and MD-MRS (37 °C, 110 rpm).

**Table 3**  
Growth curve characteristics of *L. rhamnosus* in MRS and MD-MRS.

	MRS	MD-MRS
Growth Curve Features	Lag phase until 5 h and exponential till 16th h and stationary phase from 17th h	Lag phase until 5th h followed by steady stationary phase
Increase in log CFU/mL	3.16	0.69
$K_{max}$	1.65	0.202
Td	0.6 min	4.95 min

a maltodextrin ratio of 1:1 and an inlet feed rate of 40 mL/min. The  $R^2$  value of the master model was 0.86 whereas the predictive model had an  $R^2$  value of 0.71. The optimized coded equation (–1,1) of the predictive model is presented in equation (4).

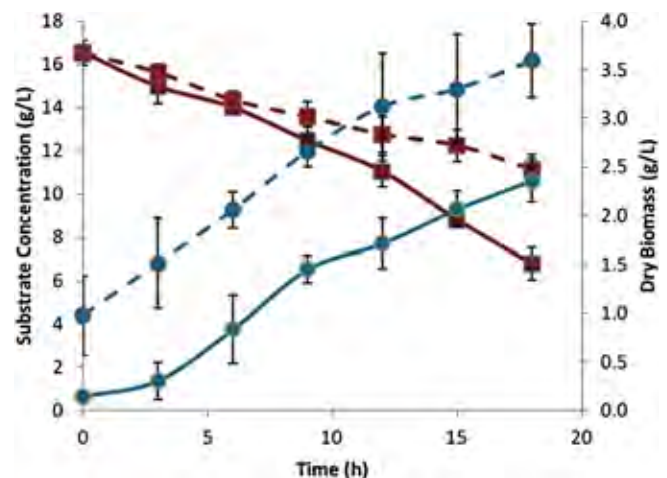
$$\% \text{ recovery} = 33.331 - 3.593 * \text{maltodextrin ratio} - 7.236 * \text{feed rate} + 4.63 * \text{maltodextrin ratio} * \text{feed rate} \quad (4)$$

### 3.4. % survival

The response plot for the % survival (in terms of log CFU/mL) for the combined probiotics is presented in Fig. 8. As expected, the processing temperature had a major effect on the probiotics' survival. The survival dropped down to almost 55% (9.5–5 log CFU/mL) when the inlet temperature was raised to 130 °C. The grid on top indicates the optimized conditions with respect to maximum % survival (81.17%) at the input conditions of 100 °C inlet temperature, a maltodextrin ratio of 1:1 and an inlet feed rate of 40 mL/min.

The  $R^2$  value of the master model was 0.96 whereas the optimized predictive model had an  $R^2$  value of 0.91. The optimized coded (–1, 1) equation of the predictive model for the probiotic survival is presented in equation (5).

$$\% \text{ survival} = 67.933 - 8.98 * \text{inlet temp} - 2.681 * \text{maltodextrin ratio} + 3.124 * \text{feed rate} + 4.863 * \text{inlet temp} * \text{maltodextrin ratio} + 5.840 * \text{maltodextrin ratio} * \text{feed rate} \quad (5)$$



**Fig. 3.** Changes in dry biomass and substrate over time for *L. acidophilus*.

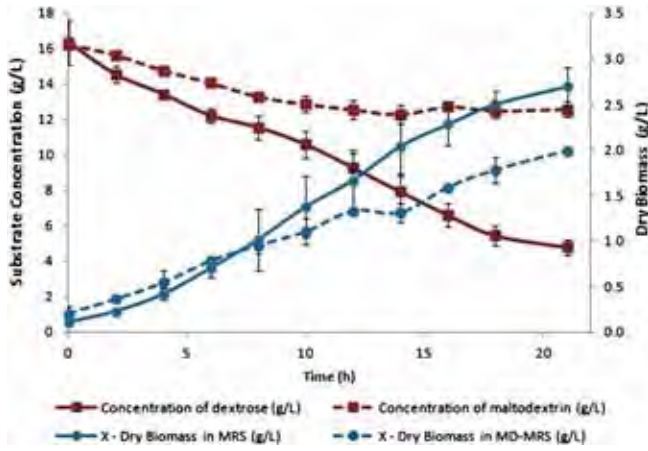


Fig. 4. Changes in dry biomass and substrate over time for *L. rhamnosus*.

3.5. Color

The grid on top of the response plot for color, as presented in Fig. 9, indicates the optimized conditions with respect to maximum

color difference ( $\Delta E$  57.210) at conditions of 100 °C inlet temperature, maltodextrin ratio of 1:1 and inlet feed rate of 40 mL/min. The  $R^2$  value of the master model was 0.91 whereas the optimized predictive model had an  $R^2$  value of 0.87. The optimized predictive model (-1, 1 coded values) for powder color is presented in equation (6).

$$\begin{aligned} \text{Color} = & 53.283 - 0.005 \cdot \text{inlet temp} \\ & - 2.068 \cdot \text{maltodextrin ratio} + 0.114 \cdot \text{feed rate} \\ & - 0.689 \cdot \text{inlet temp} \cdot \text{maltodextrin ratio} \\ & + 0.849 \cdot \text{inlet temp} \cdot \text{feed rate} \\ & + 1.808 \cdot \text{maltodextrin ratio} \cdot \text{maltodextrin ratio} \quad (6) \end{aligned}$$

Overall optimized conditions were determined at 100 °C inlet temperature, maltodextrin ratio of 1:1 and feed flow rate of 40 mL/min, where the maximum output dependent variables were obtained with 48.79% recovery, 87.17% survival and 57.21 color change ( $\Delta E$ ). The overall desirability of the model was 91.15%.

4. Discussions

In the substrate utilization analysis (Figs. 3 and 4) it was made evident that dextrose is better metabolized by the bacteria than

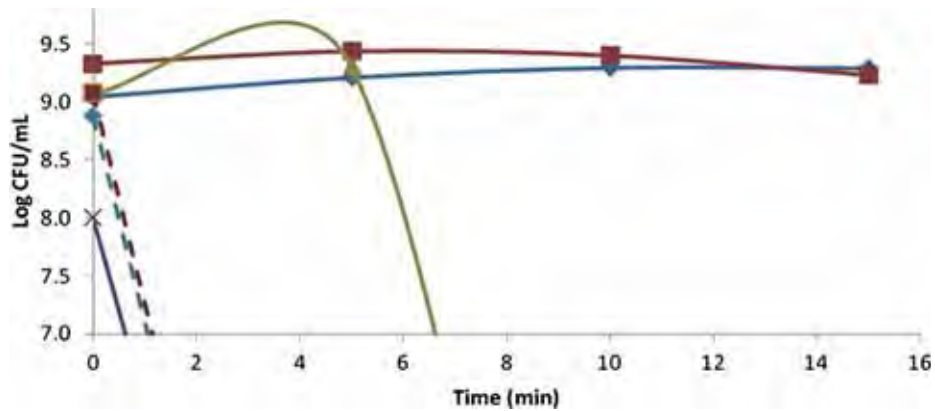


Fig. 5. Sub-lethal temperature-time assessment of *L. acidophilus*.

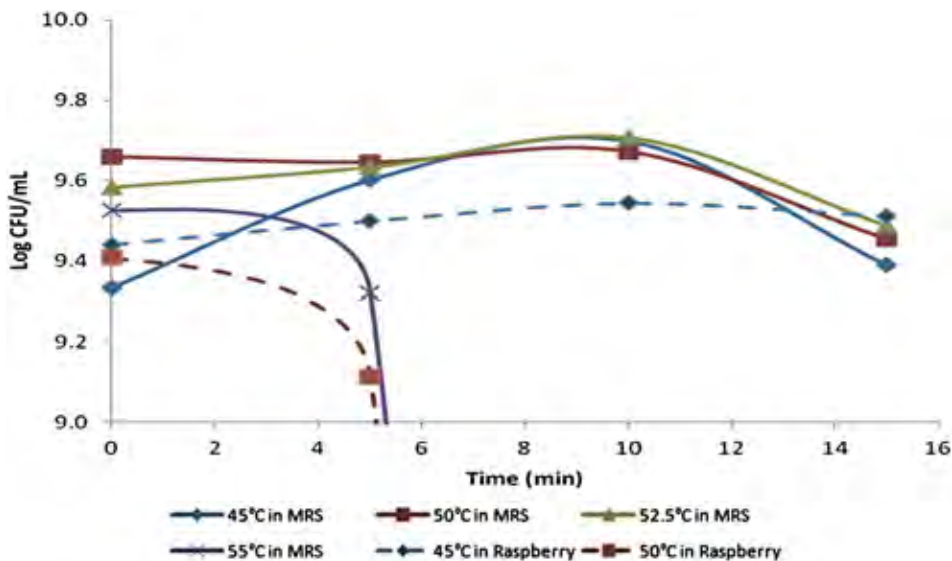


Fig. 6. Sub-lethal temperature-time assessment of *L. rhamnosus*.



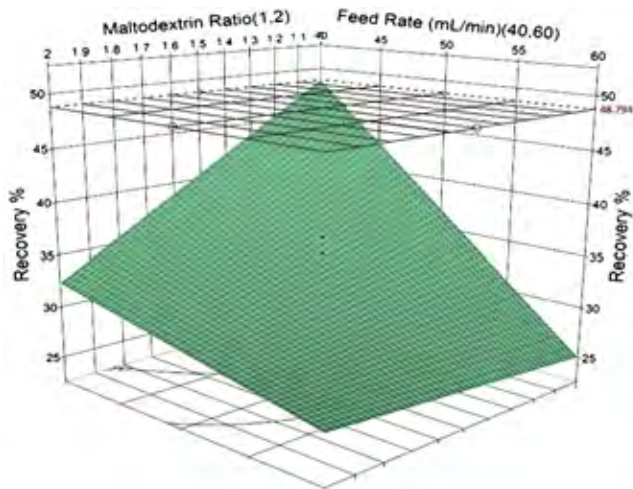


Fig. 7. Response plot of % recovery with respect to maltodextrin ratio and feed rate (mL/min).

maltodextrin. However the dry biomass at the end of the growth cycle of *L. acidophilus* in MD-MRS (3.6 g/L) was higher than that of MRS (2.4 g/L). Indeed, after the centrifugation and before drying, there was a certain amount of insoluble maltodextrin present as a residue in the pellet as a white mass (which was not seen in MRS) in *L. acidophilus* cultures (hence the dry biomass curve of MD-MRS in Fig. 3 did not start from 0). So the true weight might have been influenced by the presence of insoluble maltodextrin.

Although the organism growth was not very strong with maltodextrin it could potentially act as a prebiotic as well as a micro-encapsulating agent during spray drying. MRS acted as a better heating medium, for thermal shock pre-treatments, than raspberry because of the more complex nutrients present in MRS medium than the plain raspberry juice which has mostly sugars and fibers. It is assumed that the proteins present in the complex media may contribute to the stability of the intracellular proteins of the *Lactobacilli* during thermal shock. The sub-lethal stress induces protective mechanisms such as an alteration or reprogramming of the metabolic pathways to adjust to the new environment (Teixeira et al., 1994), thus increasing their survival during subsequent harsh treatment and improved viability during storage (O'Riordan et al., 2001; Selmer-Olsen, Birkeland, & Sørhaug, 1999; Shah & Ravula, 2000).

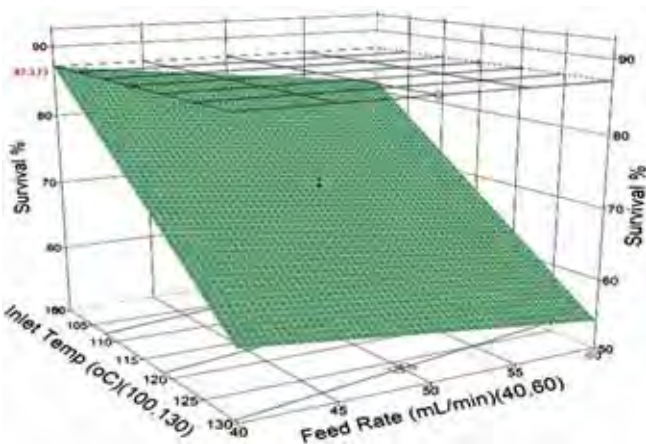


Fig. 8. Response plot for % survival with inlet temperature (°C) and feed rate (mL/min).

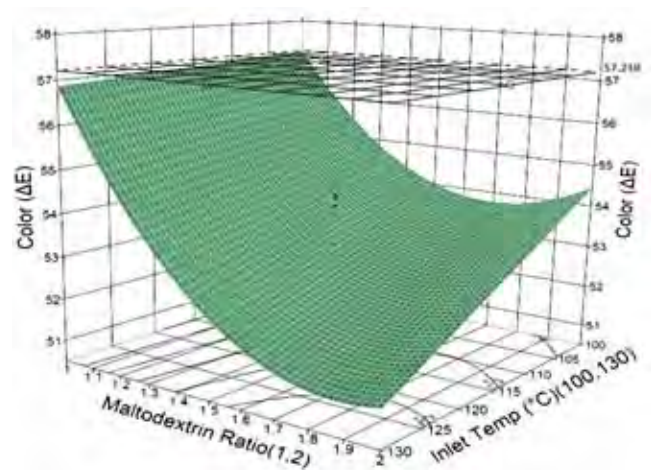


Fig. 9. Response plot of powder color with maltodextrin ratio and inlet temperature (°C) as independent variables.

Previous studies indicate that cells in the exponential growth phase are more easily adapted than in the stationary phase proving that age of cells also has a pronounced effect on the induction of thermo-tolerance (Corcoran et al., 2004; Corcoran, Ross, Fitzgerald, Dockery, & Stanton, 2006). Therefore the probiotic cultures experience both active growth and carbon depletion during the same phase of the growth cycle making them more resistant subsequently.

The walls of the spray drying chamber was layered completely with the raspberry solids by the end of spray drying due to the stickiness of the solutions which lead to major losses in product recovery. An interesting observation from the obtained data (Table 1) is that the temperature had a minimal role in powder recovery. From the response plot (Fig. 7), maltodextrin ratio and feed rate had significant effects on % recovery of produced powder from raspberry juice. As the maltodextrin ratio increased to 2, the powder recovery dropped to 33%. Reducing the maltodextrin ratio can increase the powder recovery but it will affect the cell survival of the probiotic due to reduced encapsulation efficiency. Similarly when the feed rate increased upto 60 mL/min the recovery dropped to as low as 25%. Faster feed rate does not give enough time for complete drying of the pulp thus reducing the %recovery.

On the other hand, outlet temperatures greater than 85°–90 °C are lethal for probiotics (Corcoran et al., 2004; Gardiner et al., 2000; Zamora, Carretero, & Parés, 2006) but a sub-lethal temperature pretreatment enabled cells to survive in that range with cell death occurring only after 92 °C outlet temperature in our current study. It is seen in Table 1 that under same inlet temperature conditions, a higher inlet feed rate had a lower outlet temperature and an increased survival rate. This indicates that the cell survival is mostly dependent on outlet temperature as expected from previous studies (Boza, Barbin, & Scamparini, 2004). Cellular membrane heat damage is one of the most susceptible target damage during spray drying. High temperatures during spray drying cause the cellular pores to leak the intracellular substances (Corcoran et al., 2004; Gardiner et al., 2000). The molecular nature of heat damage (above 90 °C) is not clearly known but denaturation and loss of metabolic activity of critical proteins, DNA and ribosomes are few vital events (Meng, Stanton, Fitzgerald, Daly, & Ross, 2008; Teixeira, Castro, Mohács-Farkas, & Kirby, 1997). Heat shock proteins produced during sub-lethal stress aid in protecting the probiotics during subsequent stress.

Probiotics' survival during spray drying can also attributed to the strong adherence to the carrier, which protects cells from high

**Table 4**

Color characteristics of the raspberry powder samples and their rehydrated liquid samples (1 g/9 mL of water) from all the responses.

Response	Color characteristics of powder			$\Delta E$	Color characteristics of rehydrated liquid			$\Delta E$
	L	a	b		L	a	b	
R1	58.270 ± 0.509	42.646 ± 0.769	11.496 ± 0.312	54.251 ± 0.808	24.736 ± 0.568	17.506 ± 0.260	8.953 ± 0.354	20.055 ± 0.354
R2	59.268 ± 0.195	43.108 ± 0.482	11.758 ± 0.196	54.094 ± 0.446	25.360 ± 0.392	18.095 ± 0.333	9.465 ± 0.325	20.718 ± 0.436
R3	60.242 ± 0.549	42.315 ± 0.534	11.018 ± 0.403	52.766 ± 1.646	24.913 ± 0.259	17.883 ± 0.660	9.427 ± 0.641	20.496 ± 0.919
R4	57.410 ± 0.127	44.795 ± 0.474	12.855 ± 0.021	56.719 ± 0.444	24.890 ± 0.099	19.490 ± 0.552	10.045 ± 0.247	22.268 ± 0.612
R5	60.535 ± 0.054	41.870 ± 0.156	10.372 ± 0.172	52.187 ± 0.188	24.710 ± 0.189	16.763 ± 0.443	9.047 ± 0.236	19.42 ± 0.471
R6	59.850 ± 0.786	41.288 ± 2.132	10.913 ± 0.709	52.149 ± 2.270	24.493 ± 0.238	17.970 ± 0.206	9.438 ± 0.230	20.698 ± 0.23
R7	56.298 ± 0.053	43.935 ± 0.021	12.513 ± 0.286	56.623 ± 0.106	25.238 ± 0.180	19.453 ± 0.173	9.998 ± 0.173	22.180 ± 0.212
R8	60.454 ± 0.101	42.599 ± 0.087	10.981 ± 0.217	52.689 ± 0.281	24.821 ± 0.762	17.163 ± 1.425	9.259 ± 0.900	19.876 ± 1.537
R9	60.213 ± 0.456	41.460 ± 1.866	10.813 ± 0.590	52.062 ± 1.8	24.485 ± 0.306	17.335 ± 0.306	9.218 ± 0.098	20.037 ± 0.287
R10	58.199 ± 0.225	43.976 ± 0.394	11.996 ± 0.164	55.443 ± 0.475	23.463 ± 0.293	14.674 ± 0.281	7.735 ± 0.304	17.262 ± 0.314
R11	56.650 ± 0.118	44.032 ± 0.172	12.035 ± 0.068	56.394 ± 0.218	24.557 ± 0.094	18.145 ± 0.021	9.167 ± 0.019	20.741 ± 0.013
R12	59.612 ± 0.276	42.608 ± 0.483	10.783 ± 0.165	53.313 ± 0.575	23.788 ± 0.205	16.712 ± 0.016	8.687 ± 0.099	19.381 ± 0.017
R13	59.520 ± 0.382	42.260 ± 1.068	10.977 ± 0.236	53.413 ± 0.851	24.750 ± 0.370	17.130 ± 0.122	8.953 ± 0.127	19.708 ± 0.204
R14	58.840 ± 0.255	43.120 ± 0.320	11.707 ± 0.075	54.337 ± 0.357	25.207 ± 0.193	18.457 ± 0.206	9.827 ± 0.104	21.208 ± 0.238
R15	61.310 ± 0.297	41.615 ± 0.714	10.785 ± 0.021	51.558 ± 0.747	25.125 ± 0.163	15.750 ± 0.990	8.675 ± 0.276	18.297 ± 0.986
R16	56.036 ± 0.430	44.234 ± 0.613	12.519 ± 0.210	57.082 ± 0.775	25.030 ± 0.120	19.783 ± 0.138	10.050 ± 0.138	22.521 ± 0.121
R17	60.053 ± 0.100	42.753 ± 0.514	11.093 ± 0.229	53.238 ± 0.506	24.063 ± 0.240	16.040 ± 0.115	8.480 ± 0.173	18.642 ± 0.203
R18	55.903 ± 0.818	45.167 ± 1.000	13.480 ± 0.696	58.018 ± 1.407	25.273 ± 0.542	19.727 ± 0.152	10.140 ± 0.125	22.485 ± 0.215
R19	59.413 ± 0.141	42.332 ± 0.394	10.545 ± 0.191	53.165 ± 0.268	24.212 ± 0.021	16.775 ± 0.016	8.783 ± 0.094	19.4 ± 0.024
R20	56.750 ± 0.173	44.860 ± 0.349	12.928 ± 0.278	54.337 ± 0.357	23.098 ± 0.209	15.830 ± 0.116	8.143 ± 0.168	21.208 ± 0.238

acidic and bile conditions (Crittenden et al., 2001). Overall, maltodextrin is confirmed to serve as a good encapsulating matrix as well as a moderate prebiotic for high survival of probiotics (Cortés-Arminio, López-Malo, & Palou, 2010). The sugars present in the raspberry may also have contributed to the survival during drying since sugars act as thermoprotectants during spray drying (Carvalho et al., 2003).

The response plot of powder color (Fig. 8) shows that the maltodextrin ratio had a major effect on color ( $\Delta E$ ) as it reduced (reaching toward white or increase in whiteness) with an increasing concentration of maltodextrin. On the other hand, the rehydrated liquid (Table 4), produced from the powder for all the three MD ratios had similar bright red color. So, in the case where consumption of the raspberry powder is via rehydration, then maltodextrin ratio can be increased further as the rehydrated liquid is more or less bright red in color irrespective of the MD concentration (Table 4).

## 5. Conclusions

Dextrose was a better carbon source than maltodextrin in support of the growth of the *Lactobacilli* (*L. acidophilus* and *L. rhamnosus*) chosen in the current study. MRS medium acted as a better heating medium than raspberry juice during the sub-lethal heat shock pre-treatment. Microorganisms were able to withstand up to 50 °C (for *L. acidophilus*) and 52.5 °C (for *L. rhamnosus*) in MRS as heating medium whereas both the microorganisms were killed at 45 °C in raspberry juice as the heating medium.

Maltodextrin ratio and inlet feed rate have a major effect on recovery and the optimization equation followed a linear two factor model. Inlet temperature had a profound effect on % survival during spray drying. Although in the two factor linear optimization equation, the other independent variables inlet feed rate and maltodextrin ratio and also their products were significant. Maltodextrin ratio had a major effect on color of the raspberry powder.

To the best of our knowledge there are no previous studies which optimized the microencapsulation of probiotics in fruit powder by spray drying. The overall desirability of the optimized model was 91.15% which is quite high for a biological system on which there is minimal process control. In conclusion, the choice of the microencapsulating agent and its interaction between the

protein-carbohydrate-probiotics formulation should be studied further to ensure minimal toxicity and better bioavailability. Health benefits must be validated in the presence of food matrix dosage rather than simply with isolated pure culture.

## Acknowledgments

The authors would like to acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada (NSERC).

## References

- Azcarate-Peril, M. A., Tallon, R., & Klaenhammer, T. R. (2009). Temporal gene expression and probiotic attributes of *Lactobacillus acidophilus* during growth in milk. *Journal of Dairy Science*, 92, 870–886.
- Boylston, T. D., Vinderola, C. G., Ghoddusi, H. B., & Reinheimer, J. A. (2004). Incorporation of bifidobacteria into cheeses: challenges and rewards. *International Dairy Journal*, 14, 375–387.
- Boza, Y., Barbin, D., & Scamparini, A. R. P. (2004). Effect of spray-drying on the quality of encapsulated cells of *Beijerinckia* sp. *Process Biochemistry*, 39, 1275–1284.
- Broadbent, J. R., & Lin, C. (1999). Effect of heat shock or cold shock treatment on the resistance of *Lactococcus lactis* to freezing and lyophilization. *Cryobiology*, 39, 88–102.
- Burgain, J., Gaiani, C., Linder, M., & Scher, J. (2011). Encapsulation of probiotic living cells: from laboratory scale to industrial applications. *Journal of Food Engineering*, 104, 467–483.
- Carvalho, A. S., Silva, J., Ho, P., Teixeira, P., Malcata, F. X., & Gibbs, P. (2003). Effects of Addition of Sucrose and Salt, and of Starvation upon thermotolerance and survival during storage of freeze-dried *Lactobacillus delbrueckii* ssp. *bulgaricus*. *Journal of Food Science*, 68, 2538–2541.
- Chávez, B. E., & Ledebor, A. M. (2007). Drying of probiotics: optimization of formulation and process to enhance storage survival. *Drying Technology*, 25, 1193–1201.
- Chegini, G. R., & Ghobadian, B. (2005). Effect of spray-drying conditions on physical properties of orange juice powder. *Drying Technology*, 23, 657–668.
- Chiou, D., & Langrish, T. A. G. (2007). Development and characterisation of novel nutraceuticals with spray drying technology. *Journal of Food Engineering*, 82, 84–91.
- Corcoran, B. M., Ross, R. P., Fitzgerald, G. F., Dockery, P., & Stanton, C. (2006). Enhanced survival of GroESL-overproducing *Lactobacillus paracasei* NFBC 338 under stressful conditions induced by drying. *Applied and Environmental Microbiology*, 72, 5104–5107.
- Corcoran, B. M., Ross, R. P., Fitzgerald, G. F., & Stanton, C. (2004). Comparative survival of probiotic lactobacilli spray-dried in the presence of prebiotic substances. *Journal of Applied Microbiology*, 96, 1024–1039.
- Cortés-Arminio, C., López-Malo, A., Palou, E. (2010) Agave juice as an agent for probiotic encapsulation by spray drying. In *17th World Congress of International Commission of Agricultural and Biosystems Engineering conference proceedings, Quebec City*.



- Crittenden, R., Laitila, A., Forssell, P., Maöttö, J., Saarela, M., Mattila-Sandholm, T., et al. (2001). Adhesion of *bifidobacteria* to granular starch and its implications in probiotic technologies. *Applied and Environmental Microbiology*, 67, 3469–3475.
- Dabbah, R., Moats, W. A., & Mattick, J. F. (1969). Factors affecting resistance to heat and recovery of heat-injured bacteria. *Journal of Dairy Science*, 52, 608–614. [http://dx.doi.org/10.3168/jds.S0022-0302\(69\)86615-4](http://dx.doi.org/10.3168/jds.S0022-0302(69)86615-4).
- Desmond, C., Stanton, C., Fitzgerald, G. F., Collins, K., & Paul Ross, R. (2001). Environmental adaptation of probiotic *Lactobacilli* towards improvement of performance during spray drying. *International Dairy Journal*, 11, 801–808.
- Famularo, G., de Simone, C., Matteuzzi, D., & Pirovano, F. (1999). Traditional and high potency probiotic preparations for oral bacteriotherapy. *Bio Drugs*, 12, 455–470.
- Gardiner, G. E., Bouchier, P., O'Sullivan, E., Kelly, J., Kevin Collins, J., Fitzgerald, G., et al. (2002). A spray-dried culture for probiotic Cheddar cheese manufacture. *International Dairy Journal*, 12, 749–756.
- Gardiner, G. E., O'Sullivan, E., Kelly, J., Auty, M. A. E., Fitzgerald, G. F., Collins, J. K., et al. (2000). Comparative survival rates of human-derived probiotic *Lactobacillus paracasei* and *L. salivarius* strains during heat treatment and spray drying. *Applied and Environmental Microbiology*, 66, 2605–2612.
- Hamilton-Miller, J. M. T., & Shah, S. (2002). Deficiencies in microbiological quality and labelling of probiotic supplements. *International Journal of Food Microbiology*, 72, 175–176.
- Ishibashi, N., & Shimamura, S. (1993). *Bifidobacteria*: research and development in Japan. *Food Technology*, 47, 126–135.
- Kailasapathy, K., & Rybka, S. (1997). *L. acidophilus* and *Bifidobacterium* spp. – their therapeutic potential and survival in yogurt. *Australian Journal of Dairy Technology*, 52, 28–35.
- Klaenhammer, T. R. (2000). Probiotic bacteria: today and tomorrow. *Journal of Nutrition*, 130, 415S–416S.
- Meng, X. C., Stanton, C., Fitzgerald, G. F., Daly, C., & Ross, R. P. (2008). Anhydrobiotics: the challenges of drying probiotic cultures. *Food Chemistry*, 106, 1406–1416.
- O'Riordan, K., Andrews, D., Buckle, K., & Conway, P. (2001). Evaluation of microencapsulation of a *Bifidobacterium* strain with starch as an approach to prolonging viability during storage. *Journal of Applied Microbiology*, 91, 1059–1066.
- Porubcan, R. S., & Sellers, R. L. (1975). *Stabilized dry cultures of lactic acid producing bacteria*. U.S. Patent 3,897,307, July 9.
- Ranadheera, R. D. C. S., Baines, S. K., & Adams, M. C. (2010). Importance of food in probiotic efficacy. *Food Research International*, 43, 1–7.
- Reid, G. (2002). The role of Cranberry and probiotics in intestinal and urogenital tract health. *Critical Reviews in Food Science and Nutrition*, 42, 293–300.
- Sadasivam, S., & Manickam, A. (2005). *Biochemical methods*. New Delhi: New Age International Publishers.
- Sanders, M. E. (1999). Probiotics. *Food Technology*, 53, 67–75.
- Selmer-Olsen, E., Birkeland, S. E., & Sørhaug, T. (1999). Effect of protective solutes on leakage from and survival of immobilized *Lactobacillus* subjected to drying, storage and rehydration. *Journal of Applied Microbiology*, 87, 429–437.
- Shah, N. P., & Ravula, R. R. (2000). Microencapsulation of probiotic bacteria and their survival in frozen fermented dairy desserts. *Australian Journal of Dairy Technology*, 55, 139–144.
- Teixeira, P., Castro, H., & Kirby, R. (1994). Inducible thermotolerance in *Lactobacillus bulgaricus*. *Letters in Applied Microbiology*, 18, 218–221.
- Teixeira, P., Castro, H., & Kirby, R. (1995). Spray drying as a method for preparing concentrated cultures of *Lactobacillus bulgaricus*. *Journal of Applied Bacteriology*, 78, 456–462.
- Teixeira, P. C., Castro, M. H., Malcata, F. X., & Kirby, R. M. (1995). Survival of *Lactobacillus delbrueckii* ssp. *bulgaricus* following spray drying. *Journal of Dairy Science*, 78, 1025–1031.
- Teixeira, P., Castro, H., Mohácsi-Farkas, C., & Kirby, R. (1997). Identification of sites of injury in *Lactobacillus bulgaricus* during heat stress. *Journal of Applied Microbiology*, 83, 219–226.
- Timmerman, H. M., Koning, C. J. M., Mulder, L., Rombouts, F. M., & Beynen, A. C. (2004). Monostrain, multistrain and multispecies probiotics – a comparison of functionality and efficacy. *International Journal of Food Microbiology*, 96, 219–233.
- Verdenelli, M., Ghelfi, F., Silvi, S., Orpianesi, C., Cecchini, C., & Cresci, A. (2009). Probiotic properties of *Lactobacillus rhamnosus* & *Lactobacillus paracasei* isolated from human faeces. *European Journal of Nutrition*, 48, 355–363.
- Zamora, L. M., Carretero, C., & Parés, D. (2006). Comparative survival rates of lactic acid bacteria isolated from blood, following spray-drying and freeze-drying. *Food Science and Technology International*, 12, 77–84.

Article

## Soybean Hydrophobic Protein Response to External Electric Field: A Molecular Modeling Approach

Ashutosh Singh \*, Valérie Orsat and Vijaya Raghavan

Department of Bioresource Engineering, McGill University, 21,111 Rue Lakeshore, Ste-anne-de-bellevue, QC, H9X 3V9, Canada; E-Mails: valerie.orsat@mcgill.ca (V.O.); vijaya.raghavan@mcgill.ca (V.R.)

\* Author to whom correspondence should be addressed; E-Mail: ashutosh.singh@mail.mcgill.ca; Tel.: +1-514-398-7773; Fax: +1-514-398-8387.

Received: 31 December 2012; in revised form: 30 January 2013 / Accepted: 4 February 2013 /

Published: 11 February 2013

---

**Abstract:** The molecular dynamic (MD) modeling approach was applied to evaluate the effect of an external electric field on soybean hydrophobic protein and surface properties. Nominal electric field strengths of 0.002 V/nm and 0.004 V/nm had no major effect on the structure and surface properties of the protein isolate but the higher electric field strength of 3 V/nm significantly affected the protein conformation and solvent accessible surface area. The response of protein isolate to various external field stresses demonstrated that it is necessary to gain insight into protein dynamics under electromagnetic fields in order to be able to develop the techniques utilizing them for food processing and other biological applications.

**Keywords:** molecular dynamic modeling; soybean hydrophobic protein; electromagnetic fields

---

### 1. Introduction

The field of molecular dynamic (MD) modeling is concerned with the atomic and molecular interactions that take place within the physical system that governs its microscopic and macroscopic behavior. It was developed as a tool to numerically simulate simplistic models in statistical mechanics [1], but advancement in computer technology paved the path for the development of MD modeling techniques to understand the structural and dynamic properties of biomolecules. The technique has been widely used in the field of molecular biology and pharmaceutical sciences, and it has led to the development of a wide range of efficient and accurate drug systems based on the structural and dynamic data obtained from the simulation. However, its use in food process engineering has been negligible; its

development in and application to modern sophisticated unit operations in food manufacturing is essential to evaluate the changes in the organoleptic and nutritional quality of the final food product. Current consumer preference is for high nutritional food products, manufactured through the application of energy efficient processes, which has challenged industries and researchers to develop and utilize newer techniques such as microwave or pulsed electric field and radio frequency, alone or in combination with traditional unit operations [2,3].

The development and optimization of food processes depends on mechanical limitations and the needs stated by the consumers in terms of the final product. In general, the quality of the final product is estimated by quantifying its components such as proteins, starch, mineral, vitamins and other constituents. However, the quality of the constituents themselves is rarely considered as a parameter for the optimization process. Application of unit operations such as drying, freezing, filtration, *etc.* not only negatively impacts the quantity of the food constituents but also have a significant effect on their quality. Proteins play a fundamental role for humans and are an important constituent of food. Processing methods and parameters have diverse effects on protein quality and quantity. Protein function depends on its structure and any change due to the presence of external stresses—such as heat, pH, chemicals and electromagnetic field—can lead to its denaturation leading to nonfunctional characteristics. The effect of an electromagnetic field on protein conformation has been extensively studied [4–8], but it is challenging to quantify the effect, as the alteration to conformation takes place in a very short time span (nanoscale) [9]. It is our understanding that in order to gain the insight into the mechanism of the interaction between electromagnetic fields and food constituents, it is necessary to understand the influence at atomistic and molecular level.

We have used a MD modeling approach to analyze the effect of a high electric field on the conformation of soybean hydrophobic protein (SHP), which is part of the soybean protein isolate. Soybean protein (SP) is one of the most extensively studied food proteins. Protein isolates from soybean are prepared from defatted soybean flour using a water extraction process under mild alkaline conditions [8]. SP has excellent functional and nutritional properties [10,11], but conventional thermal treatment processes used in food industries have undesirable effects on their solubility and water absorption characteristics [8,12]. Several novel processing techniques apply electromagnetic waves such as microwave and radio frequency or high electric field to achieve superior food quality [2]. In 2011, Xiang *et al.* [8] studied the structural modification induced by the application of PEF on soybean protein isolate. They reported that the application of an electric field intensity of 20–25 kV/cm and pulse number varying from 30–120 induced changes in the microenvironment of the soy protein's tryptophan residues from a less polar to a more polar environment, they indicated that these changes in the polarity was brought by partial denaturation of the protein under PEF treatment.

Hence in this study emphasis, has been put on quantification of conformational changes using root mean square deviation, radius of gyration and total dipole moment data and its effect on surface hydrophobicity and hydrophilicity of SHP.

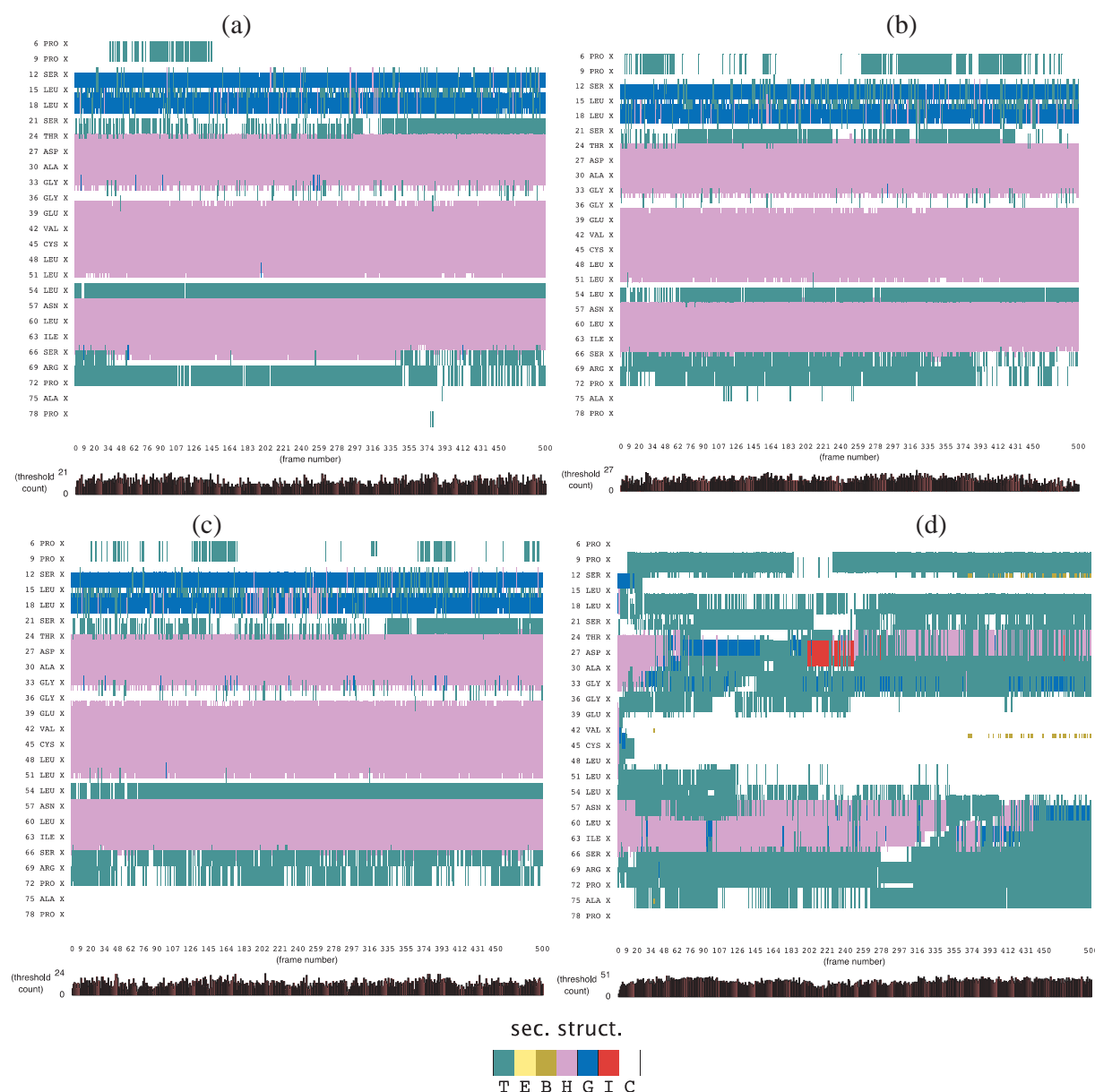
## 2. Results and Discussion

### 2.1. Secondary Structure Analysis

The effect of an external electric field on secondary structures of SHP was evaluated using the STRIDE algorithm. Figure 1 demonstrates that application of electric field strengths 0.002 V/nm and

0.004 V/nm had no major effect on the helical region of SHP, but minor changes can be observed on turns and coils, as compared to no electric field. Application of a 3V/nm electric field severely disrupts the helical region of the protein, as compared to 0.002 and 0.004 V/nm. It is interesting to observe that the application of 3V/nm lead to a transition between  $\alpha$ -helix and  $\pi$ -helix, and similar transitions were observed by Budi *et al.*, (2005) [17] in their study on the effect of an electric field on Insulin protein Chain-B conformation. They stated that this transition is an integral part of protein's dynamic property and plays an important role in defining its function [17].

**Figure 1.** STRIDE showing evolution of secondary structure of soybean hydrophobic protein (SHP) (a) without electric field (b) 0.002 V/nm (c) 0.004 V/nm and (d) 3 V/nm external electric field.



Note: Color code: Magneta denotes  $\alpha$ -helix, red denotes  $\pi$ -helix, cyan denotes turn, blue denotes 3–10 helix and white denotes coil

## 2.2. Dipole Moment Distribution

Proteins, due to their secondary structure conformations such as helices, sheets, turns, coils, *etc.* possess electric dipole moment and application of external electric field induces an alignment change in them with respect to the direction of the applied field [18]. The electric dipole of a protein is represented as:

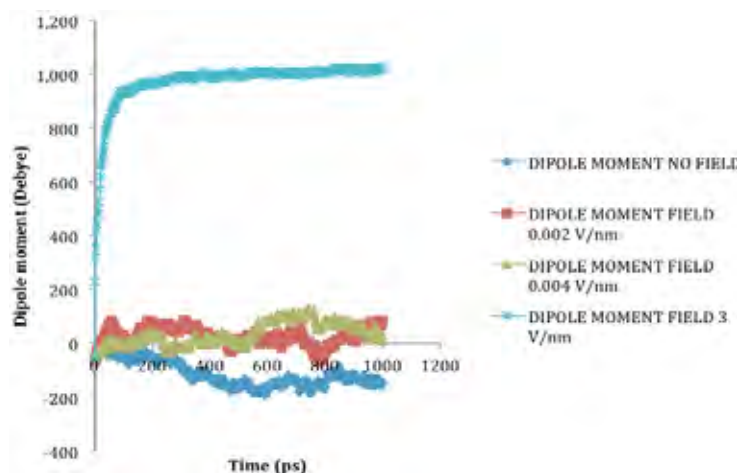
$$\vec{d} = \sum_{i=1}^N q_i(i) \vec{r}_i \quad (\text{Equation 1})$$

Where  $\vec{d}$  is the dipole,  $q_i$  is the charge of the atom  $i$ ,  $\vec{r}_i$  is the directional vector of each atom and  $N$  is the number of atoms. When an external electric field  $E$  is applied the protein orients itself in the direction of the field and in our study, the electric field was applied in the  $x$ -axis. Depending on the strength of the applied field, unfolding or re-orientation of protein is observed. From Figure 2 and Table 1 it can be noted that application of an external electric field changed the total dipole moment of SHP under MD simulation. The total dipole moment for applied field strength of 3V/nm increased until 1000 Debye in approximately 200 ps and remained constant thereafter because the protein completely unfolded and all the major secondary structures were lost (Figure 5). From Figures 3 and 4 it can also be observed that at the end of the simulation, the helices of SPI were realigned in the direction of the applied electric field, but no preferential alignment of helices were observed in the reference (no field), confirming that under an external electric field, proteins orient themselves in the direction of the field [9,15,17].

**Table 1.** RMSD, Radius of Gyration and Total Dipole moment of the SHP protein Backbone, averaged over 1ns of simulation time.

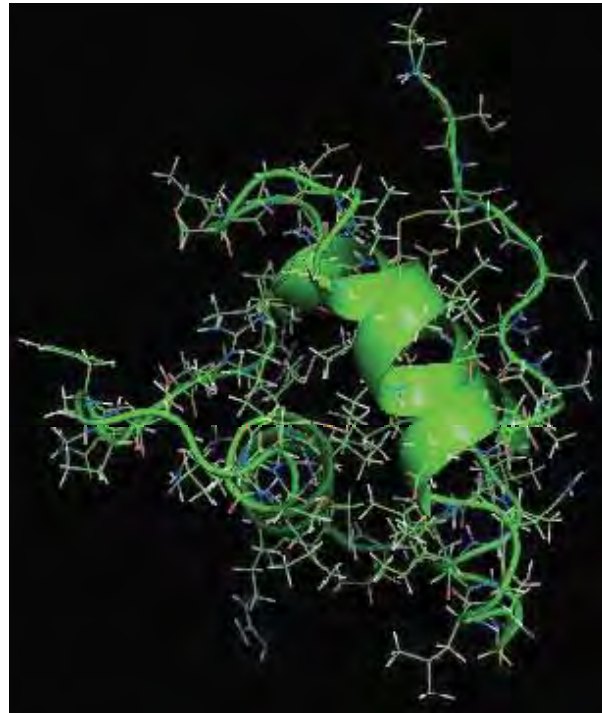
Molecule	Electric field strength (V/nm)	RMSD average (nm)	Rg average (nm)	Total Dipole moment (Debye)
SHP (1HYP)	0	0.122 ± 0.022	1.150 ± 0.024	-121 ± 45.311
SHP (1HYP)	0.002	0.114 ± 0.016	1.163 ± 0.037	27 ± 31.444
SHP (1HYP)	0.004	0.119 ± 0.021	1.150 ± 0.021	38 ± 42.460
SHP (1HYP)	3	0.803 ± 0.060	1.435 ± 0.043	971 ± 102.666

**Figure 2.** Total dipole moment of SHP under the influence of an external electric field.

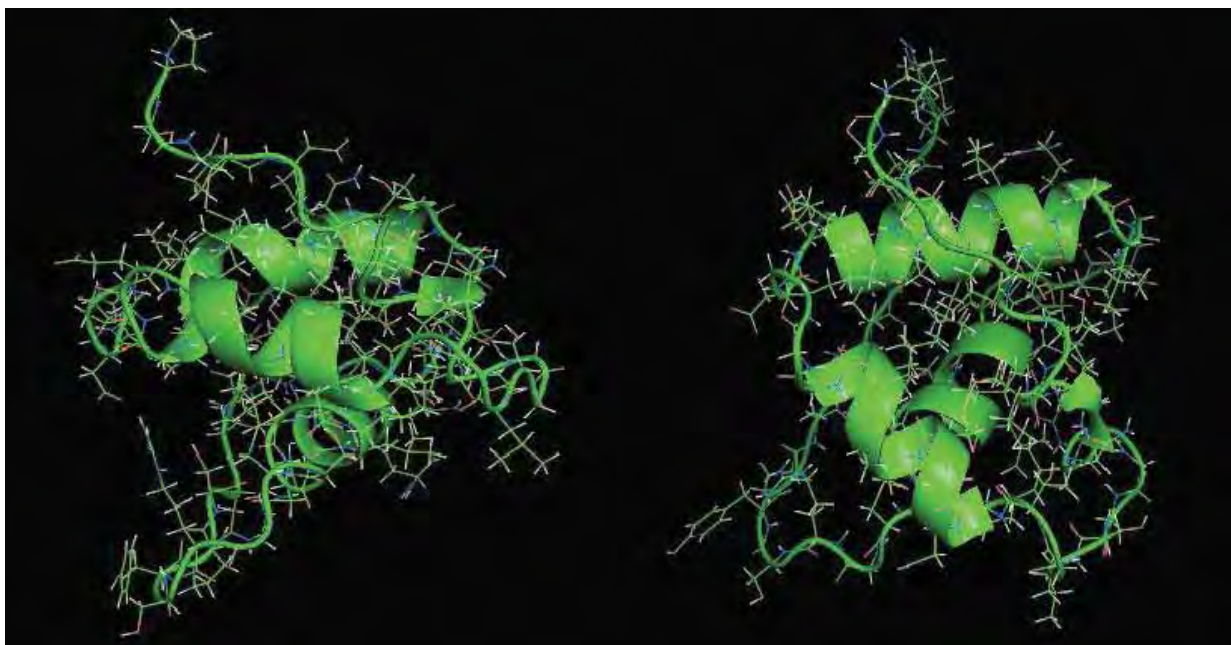




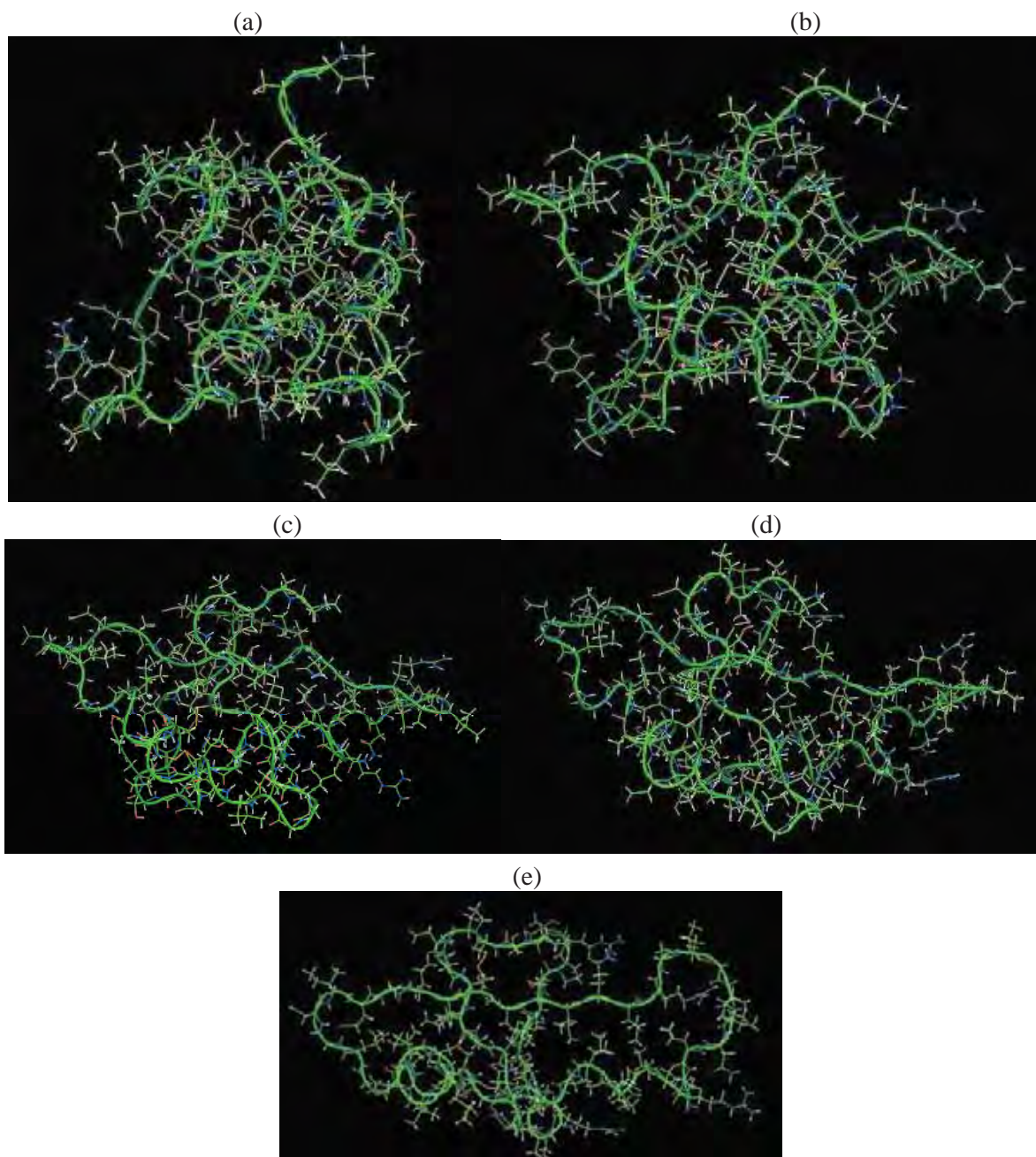
**Figure 3.** SHP protein under no external electric field.



**Figure 4.** SHP protein snapshot at the end of the MD simulation under the influence of 0.002 V/nm and 0.004 V/nm.



**Figure 5.** SHP snapshots at the beginning (a) and at different time spans (b), (c), (d) and at the end of the MD simulation (e) with electric field 3 V/nm.



### 2.3. Root Mean Square Deviation (RMSD):

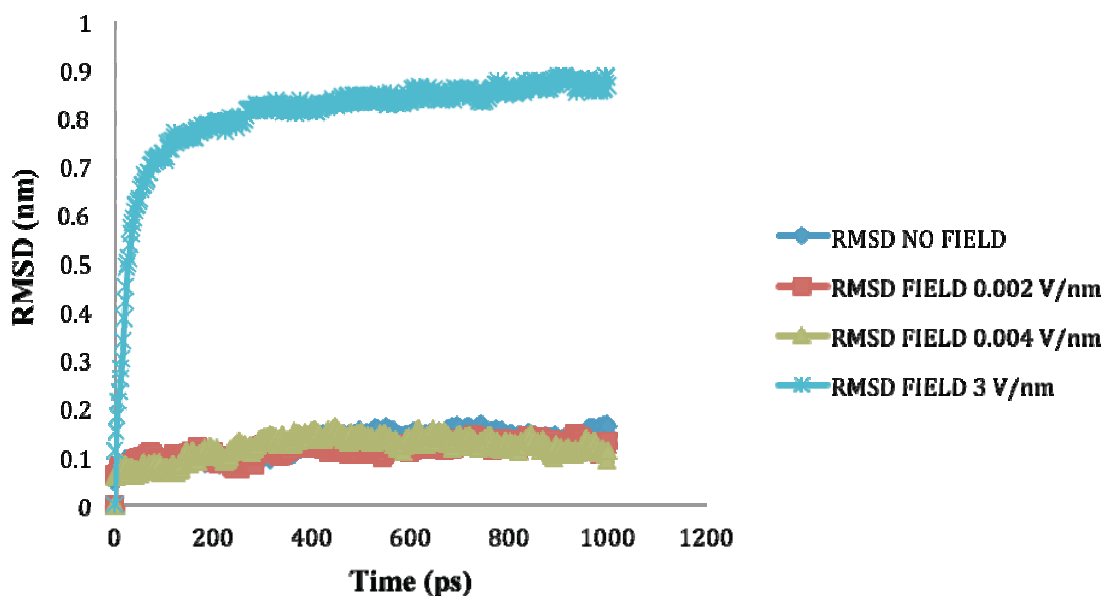
The RMSD of a protein during simulation provides a quantitative measure of any structural variation. RMSD is calculated by comparing the structure of simulated protein during simulation with the reference [19]. RMSD of a protein is represented as:

$$RMSD = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N |r_{final}(i) - r_{initial}(i)|^2} \quad (\text{Equation 2})$$

Where  $r_{final}(i)$  is the final coordinates of an atom  $i$ , and  $r_{initial}(i)$  is the initial coordinate of the atom  $i$ , and  $N$  is the number of atoms.

The average value of RMSD for 1ns simulation is shown in Table 1. It can be assessed from Table 1 and Figure 6 that as the application of an electric field strength of 0.002 V/nm and 0.004 V/nm had no major effect on the stability of the SHP structure as their RMSD value were similar to the reference (no field), but at 3V/nm the partial and complete loss of the helical secondary structures produced a higher RMSD value.

**Figure 6.** RMSD of SHP backbone under the effect of external electric field.



#### 2.4. Radius of Gyration (Rg)

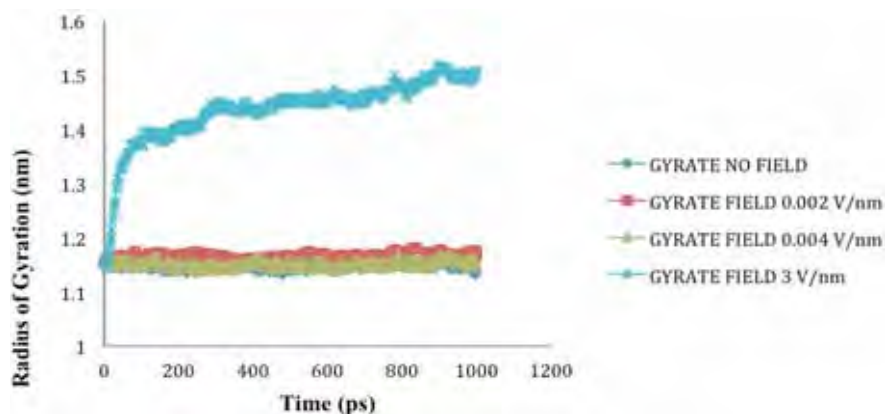
The Rg of the protein quantifies the distribution of the atoms in the space relative to their center of mass; it provides an understanding of the changes in shape and size of the protein under the influence of external stresses such as an electric field [17, 20]. It can be calculated using Equation 3.

$$Rg = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N |r(i) - r_{center}|^2} \quad (\text{Equation 3})$$

Where  $r(i)$  is the coordinates of an atom  $i$  and  $r_{center}$  is the coordinates of the protein's center of mass,  $N$  is the number of atoms.

The radius of gyration (Rg) values for SHP under the influence of electric field strength 0.002 V/nm and 0.004 V/nm did not vary much with respect to the reference, because no significant change to the structure was observed, but as the electric field of 3 V/nm was applied the Rg value obtained was higher compared to the reference and other applied intensities because of the loss of secondary structures and resulting protein unfolding (Figure 7).

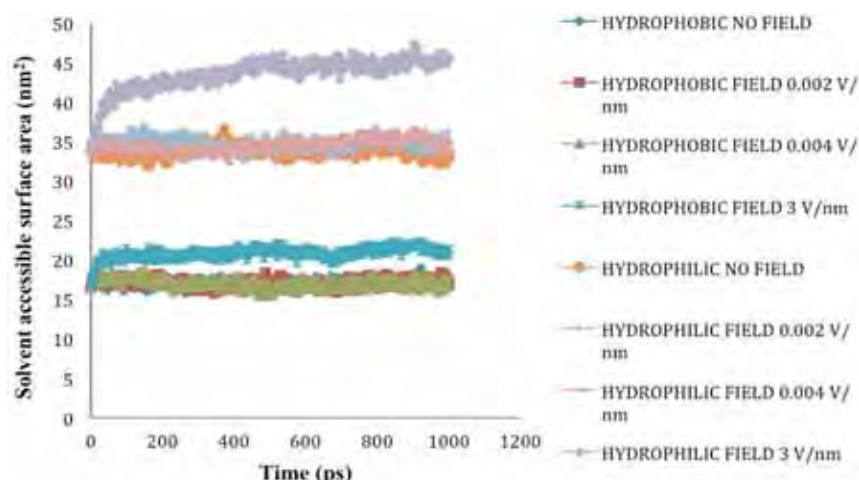
**Figure 7.** SHP radius of gyration ( $R_g$ ) in the absence and presence of various external electric fields.



### 2.5. Solvent Accessible Surface Area

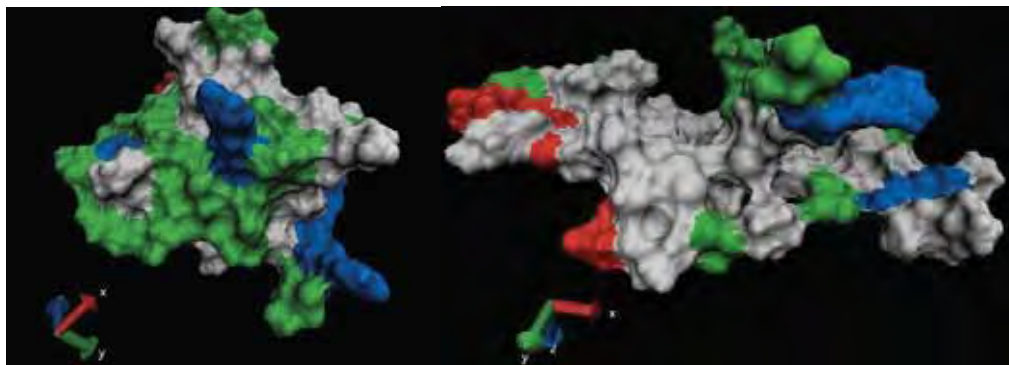
SASA denotes the area that is available on the protein surface for interaction with solvents and other molecules of similar sizes. Under cellular conditions, the hydrophobic residues of a protein are arranged in a manner to minimize exposure to solvents. However, when a protein is subjected to external stress such as thermal and high electric field, it undergoes conformational changes, which lead to increase in exposure of hydrophobic residues to water and other solvents. In our study, application of electric field stress of 0.002 V/nm and 0.004 V/nm had no noticeable change on SASA until the electric field intensity of 3V/nm was applied. Due to unfolding of SPI under this extreme stress, the hydrophobic and hydrophilic residues were exposed on the surface increasing the SASA (Figure 8–9). In 2011, Xiang *et al.* [8] reported that application of pulsed electric field treatment of 25 kV/cm and 120 pulses on soybean protein isolate (SPI) showed an 18% increase in surface hydrophobicity, they attributed these changes to denaturation of the protein. In the present study, the protein was subjected to external electric field in an aqueous system at neutral pH and the simulation was conducted at constant temperature and pH. Application of pulsed electric field or static high electric field to a protein in an aqueous environment would lead to generation of ohmic heating, in the study conducted by Xiang *et al.* [8], the effect of temperature was considered to be minimal.

**Figure 8.** Solvent accessible surface area ( $\text{nm}^2$ ) under the influence of an external electric field.





**Figure 9.** SHP snapshot of surface properties at the beginning and end of the MD simulation under electric field intensity of 3V/nm. Note: non-polar residues (white), basic (blue), acidic (red) and polar residues (green).



The findings of the present study were in accordance to several studies conducted on molecular dynamic simulation of proteins under external electric field [15,17–18]. It is important to note that the results obtained are under the assumptions of MD simulation and more analysis is required to evaluate the interaction of external field with each and every atom under the accepted force fields. Nonetheless, the MD simulation conducted for the present study does provide information about direct influence of an external electric field on the SPH.

### 3. Experimental Section

MD simulation on SHP was performed using a classical MD algorithm as implemented in Groningen machine for chemical simulations (GROMACS) software package, version 4.5.5 from the Stockholm Center for Biomembrane Research, Stockholm, Sweden [13]. SHP consists of 75 amino acid residues, where more than 50% of the secondary structure consists mainly of helices (4 helices; 3 -  $\alpha$  helix (Val25-Gly33, Asp37-Leu51 and Leu56-Ser66) and 1–3/10 helix (Ser12-Gly19) (Figure 10) (PDB sequence detail). The SHP molecule surface contains 70% apolar atoms and the four disulphide bonds found in SHP provide stability to the protein (Figure 10). SHP starting configuration (PDB accession code 1HYP) was used for our study. All atom CHARMM27 force field was used to describe the potential energy and provide functions and parameters for every type of atom in the system [9]. The SHP configuration was enclosed in a periodic cubic water box at a distance of 10 Å from the edge of the box to satisfy the minimum image convention. The water model used was TIP3P [9] and one sodium ion was added to neutralize the system. The neutral solvated protein system was first energy minimized using steepest descent for 20000 steps and equilibrated at constant volume (NVT) and temperature (NPT) for 200 ps. Due to limited computing power available, the MD simulation was run for 1000 ps with temperature maintained at 300° K and using Berendsen thermostat and pressure was maintained at 1 bar. The constant temperature and pressure coupling was 0.1 ps and 2 ps respectively. To limit the short-range nonbonded interactions, van der Waals interaction and long-range electrostatic interactions, a cut-off of 1 nm was used. PME algorithm was used with grid spacing of 0.16 nm and the time step during the simulation was 2 fs [14]. MD simulations were run under constant electric field of 0.002 V/nm, 0.004 V/nm and 3 V/nm (Table 2). One MD simulation was run without electric



field as a reference. The values of electric field chosen cover the achievable range of 20–40 kV/cm (0.002 V/nm and 0.004 V/nm) used commonly for pulsed electric field and electrohydrodynamic processing of food products [2]. The highest value of 3 V/nm corresponds to the values where other researchers observed structural changes in other proteins such as Chignolin [9,15]. All the electric fields were applied in the  $x$  axis to the equilibrated solvated protein system.

Root mean square deviation (RMSD) and radius of gyration (Rg) of the backbone atoms were calculated to represent any conformational changes in the protein. Surface hydrophobicity and hydrophilicity of the system were also estimated to evaluate the effect of electric field on the electrical and surface properties. Changes in these parameters were used as an index of protein conformational alteration and studied using GROMACS analyzing tools. The effect of external electric field on secondary structure of the protein was characterized using STRIDE algorithm implemented in visual molecular dynamics (VMD) [16]. The snapshots of protein conformational change were taken using PYMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.

**Table 2.** Summary of parameters used in the MD simulated systems.

Protein System	Electric field strength (V/nm)	Temperature (K) and Pressure (bar)	Simulation length (ns)
SHP (1HYP)	0.002	300 K, 1 bar	1
SHP (1HYP)	0.004	300 K, 1 bar	1
SHP (1HYP)	3	300 K, 1 bar	1

**Figure 10.** FASTA sequence of SPH protein, four disulphide bonds between Cys8-Cys43, Cys14-Cys28, Cys45-Cys77 and Cys29-Cys67 (PDB accession code 1HYP).



#### 4. Conclusions

In this work, by means of MD simulation, we explored the effect of external static electric fields of 0.002 V/nm, 0.004 V/nm and 3V/nm strength on the structural stability of soybean protein isolate. Our observation indicates that under the effect of lower intensity electric field, SHP protein reorients itself in the direction of the electric field and no major effect was observed on the secondary structure of the protein, but under 3V/nm, the protein unfolded and almost all the helical structures were lost during simulation. It can also be concluded that more work with oscillating electric field at different frequencies is required in order to estimate its effect on proteins' structural and functional properties. Knowledge gained through MD simulation under external electric field will be useful to explain the effect of novel food processing techniques such as microwave, radiofrequency, pulsed electric field and electrohydrodynamics on the biochemical composition of food products.

#### Acknowledgments

The authors are grateful to NSERC (Natural Sciences and Engineering Research Council of Canada) for their financial support for this study.

## References

1. Alder, B.J.; Wainwright, T.E. Studies in molecular dynamics. I. General method. *J. Chem. Phys.* **1959**, *31*, 459–466.
2. Singh, A.; Orsat, V.; Raghavan, V. A comprehensive review on electrohydrodynamic drying and high voltage electric field in the context of food and bioprocessing. *Drying Tech.* **2012**, *30*, 1812–1820.
3. Venkatesh, M.S.; Raghavan, G.S.V. An Overview of Microwave Processing and Dielectric Properties of Agri-food Materials. *Biosys. Eng.* **2004**, *885*, 1–18.
4. Fernandez-Diaz, M.D.; Barsotti, L.; Dumay, E.; Cheftel, J.C. Effects of pulsed electric fields on ovalbumin solutions and dialyzed egg white. *J. Agri. Food Chem.* **2000**, *48*, 2332–2339.
5. Laurence, J.A.; French, P.W.; Lindner, R.A.; McKenzie, D.R. Biological effects of electromagnetic fields - Mechanisms for the effects of pulsed microwave radiation on protein conformation. *J. Theor. Bio.* **2000**, *206*, 291–298.
6. De Pomerai, D.I.; Smith, B.; Dawe, A.; North, K.; Smith, T.; Archer, D.B.; Duce, I.R.; Jones, D.; Candido, E.P.M. Microwave radiation can alter protein conformation without bulk heating. *FEBS Lett.* **2003**, *543*, 93–97.
7. Li, Y.; Chen, Z.; Mo, H. Effects of pulsed electric fields on physicochemical properties of soybean protein isolates. *LWT - Food Sci. Tech.* **2007**, *40*, 1167–1175.
8. Xiang, B.Y.; Ngadi, M.O.; Simpson, B.K.; Simpson, M.V. Pulsed electric field induced structural modification of soy protein isolate as studied by fluorescence spectroscopy. *J. Food Proc. Preserv.* **2011**, *35*, 563–570.
9. Astrakas, L.; Gousias, C.; Tzaphlidou, M. Electric field effects on chignolin conformation. *J. Appli. Phys.* **2011**, *109*, 094702.
10. Sorgentini, D.A.; J.R. Wagner, and M.C. Añón, Effects of thermal treatment of soy protein isolate on the characteristics and structure-function relationship of soluble and insoluble fractions. *J. Agri. Food Chem.* **1995**, *43*, 2471–2479.
11. Puppo, C.; Chapleau, N.; Speroni, F.; De Lamballerie-Anton, M.; Michel, F.; Añón, C.; Anton, M. Physicochemical Modifications of High-Pressure-Treated Soybean Protein Isolates. *J. Agri. Food Chem.* **2004**, *52*, 1564–1571.
12. Kinsella, J.E. Functional properties of soy proteins. *J. Am. Oil Chem. Soc.* **1979**, *56*, 242–258.
13. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J.C. GROMACS: Fast, flexible, and free. *J. Comp. Chem.* **2005**, *26*, 1701–1718.
14. Petersen, H.G. Accuracy and efficiency of the particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 3668–3679.
15. Astrakas, L.G.; Gousias, C.; Tzaphlidou, M. Structural destabilization of chignolin under the influence of oscillating electric fields. *J. Appli. Phys.* **2012**, *111*, 074702.
16. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.
17. Budi, A.; Legge, F.S.; Treutlein, H.; Yarovsky, I. Electric field effects on insulin chain-B conformation. *J. Phys. Chem. B* **2005**, *109*, 22641–22648.
18. Budi, A.; Legge, F.S.; Treutlein, H.; Yarovsky, I. Effect of frequency on insulin response to electric field stress. *J. Phys. Chem. B* **2007**, *111*, 5748–5756.

19. Budi, A.; Legge, S.; Treutlein, H.; Yarovsky, I. Effect of external stresses on protein conformation: A computer modelling study. *Eur. Biophys. J.* **2004**, *33*, 121–129.
20. Budi, A.; Legge, F.S.; Treutlein, H.; Yarovsky, I. Comparative study of insulin chain-B in isolated and monomeric environments under external stress. *J. Phys. Chem. B* **2008**, *112*, 7916–7924.

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

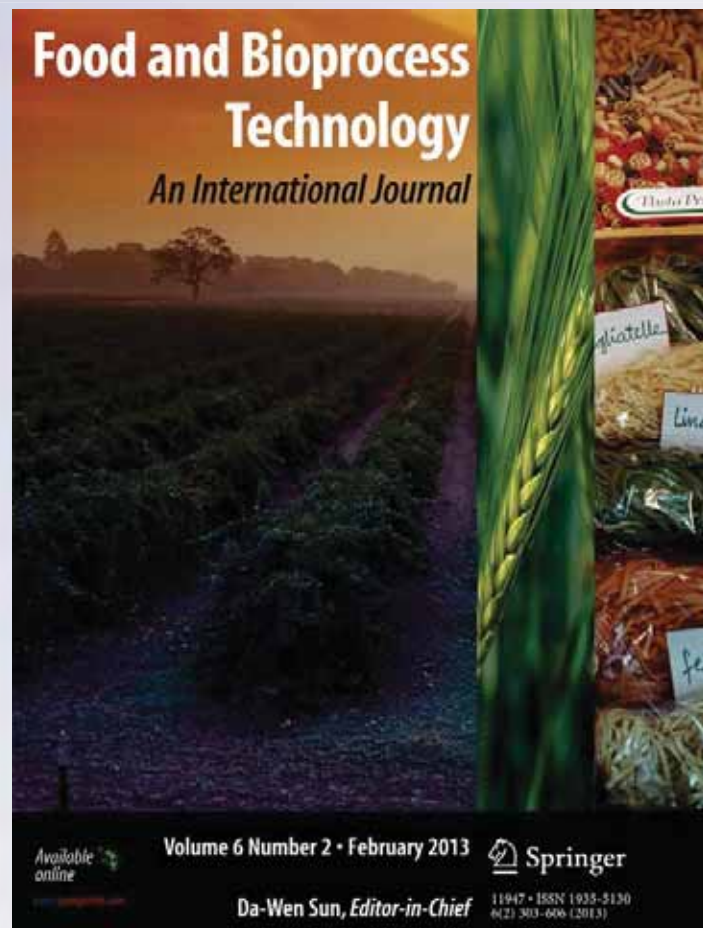
# *Optimization of Microwave-Assisted Extraction of Phenolic Antioxidants from Grape Seeds (Vitis vinifera)*

**Kiruba Krishnaswamy, Valérie Orsat,  
Yvan Gariépy & K. Thangavel**

**Food and Bioprocess Technology**  
An International Journal

ISSN 1935-5130  
Volume 6  
Number 2

Food Bioprocess Technol (2013)  
6:441-455  
DOI 10.1007/s11947-012-0800-2



**Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media, LLC. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.**



# Optimization of Microwave-Assisted Extraction of Phenolic Antioxidants from Grape Seeds (*Vitis vinifera*)

Kiruba Krishnaswamy · Valérie Orsat · Yvan Gariépy · K. Thangavel

Received: 27 June 2011 / Accepted: 1 February 2012 / Published online: 26 February 2012  
© Springer Science+Business Media, LLC 2012

**Abstract** Grape seeds (*Vitis vinifera*) are rich in phytochemicals that have antioxidant properties. The influence of independent variables such as microwave power (100, 150, and 200 W), extraction time (2, 4, and 6 min), and solvent concentration (30%, 45%, and 60% ethanol) and their interactions on total phenols and the antioxidant activity (1,1-diphenyl-2-picrylhydrazyl (DPPH) and ferric ion reducing antioxidant power (FRAP)) were determined; and the microwave-assisted extraction (MAE) process was optimized using a central composite design. The total phenols that were expressed as gallic acid equivalents (GAE), catechin equivalents (CAT), and tannic acid equivalents (TAE) were significantly influenced by the solvent concentration and the time of extraction. A numerical optimization was carried out to obtain the overall conditions for MAE of phenolic antioxidants from grape seed. The response variables were maximized for 6 min of MAE of grape seed (GS) with 32.6% ethanol at 121 W with a desirability function of 0.947. The predicted extraction yields were  $13 \pm 0.89$ ,  $21.6 \pm 1.59$ , and  $15.9 \pm 1.32$  mg GAE, CAT, and TAE, respectively per gram of GS. The predicted antioxidant activity per gram of dry weight GS was 80.9% for the inhibition of DPPH and 135  $\mu$ M ascorbic acid equivalents for FRAP test. The predicted response values were significantly correlated with the

observed ones as follows: GAE  $r=0.995$ , CAT  $r=0.990$ , TAE  $r=0.996$ , DPPH  $r=0.996$ , and FRAP  $r=0.996$ .

**Keywords** Gallic acid · Catechin · Tannic acid · Response surface methodology · Antioxidant activity · FRAP · DPPH

## Introduction

Antioxidants are substances that neutralize free radicals. Free radicals are damaging compounds in the human body that alter cell membranes, causing cell damage and inflammation, promoting abnormal cell growth including various kinds of cancer, and even causing cell death (Lee et al. 2004). Free radicals are generated naturally in the body by metabolic process and from exogenous factors like ultraviolet light, radiation, smoke, and air pollution (Kim et al. 2006). The normal aerobic respiration in cells produces natural antioxidants which are further supplemented by dietary sources like vitamins, carotenoids, flavonoids, etc. (Harman 1995). The effectiveness of the endogenous antioxidant system declines with aging. Hence it is required to incorporate antioxidative nutraceuticals in the prevention of diseases. The health-related issues with the growing population has led to the search of naturally occurring antioxidants in food to replace synthetic antioxidants, which are being restricted, due to their adverse side effects, such as carcinogenicity (Pokorn 1991).

The present regulatory regimes on waste management and disposal in the food industries has urged for an integrated approach in context of recycling/reuse and waste recovery. Recycling of fruit and vegetable waste is one of the most important means of utilizing it in a number of innovative ways yielding essentially new products meeting the requirements in human, animal, and plant nutrition as well as in the

K. Krishnaswamy · V. Orsat (✉) · Y. Gariépy  
Department of Bioresource Engineering, McGill University,  
21,111, Lakeshore Road,  
Ste-Anne-de-Bellevue, QC H9X3V9, Canada  
e-mail: valerie.orsat@mcgill.ca

K. Thangavel  
Department of Food and Agricultural Process Engineering,  
Tamil Nadu Agricultural University,  
Coimbatore 641003, India

pharmaceutical industry. Hence there is a growing interest in exploiting agricultural waste materials to obtain phytochemicals and antioxidants which will provide satisfactory solutions to both food processing industries and the society. In this context, winery waste could be an excellent source for obtaining natural antioxidants from grape pomace (Bail et al. 2008).

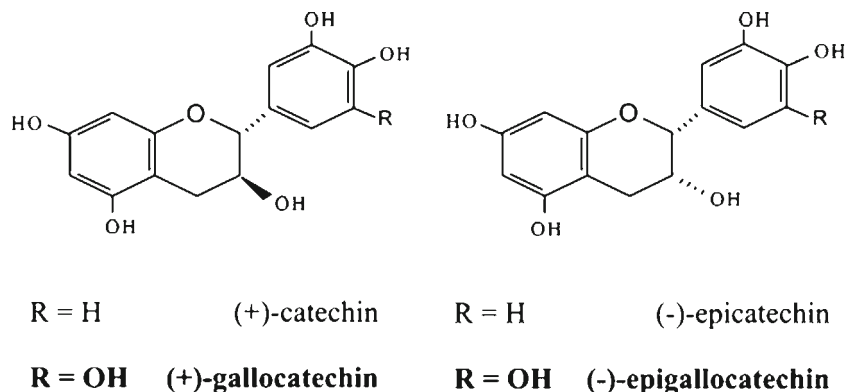
Grape (*Vitis vinifera* L.) is one of the world's most important fruit crop with a global production of around 67 million tons in 2009 (FAOSTAT 2011) and about 80% of the total production is used for wine making. The grape pomace obtained, as winery by-product, constitutes 10–20% of the weight of grapes (Garcia-Marino et al. 2006). It was reported that the grape seed present in the pomace weighs 38–52% on dry matter basis (Maier et al. 2009). Grape seed consists of 35% fiber, 29% polyphenols, 15% lipids, 3% ash, and 7% water (Mayer et al. 2008). Grape seeds are rich sources of monomeric phenolic compounds (Fig. 1) such as (+)-catechins, (–)-epicatechin, (+)-gallocatechins, (–)-epigallocatechin, and their dimeric, trimeric, and tetrameric proanthocyanidins (Baoshan Sun and Spranger 2005; Freitas et al. 1998).

Grape seed extract has been a matter of intense investigations with respect to its potential beneficial effects on human health as it possesses a broad spectrum of biological, pharmacological, and chemoprotective properties against free radicals and oxidative stress. Recent reports indicate a wide range of biological activities, e.g., antioxidant properties and radioprotective effects (Castillo et al. 2000), antibacterial activity (Jayaprakasha et al. 2003), prevention of cataract (Yamakoshi et al. 2002), antihyperglycemic effects (Pinet et al. 2004), anti-inflammatory effects (Terra et al. 2007), enhancement of postprandial lipemia (Del Bas et al. 2005), improvement of insulin sensitivity, and prevention of hypertriglyceridemia and cardiovascular disease (Al-Awwadi et al. 2005), prevention of in vitro low density lipoprotein oxidation (Meyer et al. 1997), modulation of the expression of antioxidant enzyme systems (Puiggròs et al. 2005), and protective effects against oxidative damage in mouse brain cells (Guo et al. 2007). Thus effective extraction and concentration of these phenolic bioactive compounds from the grape seed is

very important so that it can be utilized in the preparation of functional foods, dietary supplements, and in nutraceutical and pharmaceutical products (Cacace and Mazza 2003).

Microwave-assisted extraction (MAE) has been reported as a novel method for extraction of bioactive compounds. It utilizes microwave energy to cause molecular movements and rotation of liquids with permanent or induced dipoles (Sun et al. 2007). When a biological material with favorable dielectric properties (plant material along with solvent for extraction) is placed in a microwave field, the molecules try to align with the oscillating electromagnetic field either by distortion or distribution of electron cloud within the molecule or by physical rotation of the molecular dipoles which leads to rapid heating of solvent and sample matrix (Dandekar and Gaikar 2002). Thus MAE is advantageous over conventional extraction techniques with improved efficiency, reduced extraction time, rapid and volumetric heating of the absorbing medium, low solvent consumption, higher selectivity of target molecules, and high potential for automation (Pallaroni et al. 2002, Thostenson and Chou 1999, Wang et al. 2010). In conventional techniques such as heating, boiling, or refluxing, there is a loss of flavonoids due to ionization, hydrolysis, and oxidation on prolonged extraction time. MAE is the simplest, fastest, and most economical technique for extraction. The enhancement of product recovery by microwave is generally attributed to its heating effect, which occurs due to the dipole rotation of the solvent in the microwave field. This causes the solvent temperature to rise, which then increases the solubility of the compound of interest such as flavonoids (Wang et al. 2010). Microwave-assisted extraction has been used to extract many different bioactive compounds, such as 1,8 cineole, menthone, and menthol from peppermint and 1,8 cineole and camphor from rosemary leaves (Chen and Spiro 1994), essential oil from *Lippia sidoides* (Craveiro et al. 1989), alkaloid from lupin seeds (*Lupinus mutabilis*) (Ganzler et al. 1990), secoisolariciresinol diglucoside from flaxseeds (Nemes and Orsat 2010), pectin from apple pomace (Wang et al. 2007), phenolic antioxidants from peanut skin (Ballard et al. 2009), glycyrrhizic acid from licorice roots (Pan et al. 2000), curcuminoids from *Curcuma longa* (Dandekar and Gaikar

**Fig. 1** Monomeric units of proanthocyanidins present in grape seeds (Baoshan Sun and Spranger 2005)



2002), Zearalenone from wheat and corn (Pallaroni et al. 2002), anthocyanin from red raspberries (Sun et al. 2007), tea polyphenols and tea caffeine from green tea leaves (Pan et al. 2003), etc.

The optimization of microwave-assisted extraction of polyphenolic antioxidants from grape seeds was conducted in the present study with the objective of recovering maximum antioxidants. Response surface methodology (RSM) was used to evaluate the effect of multiple factors like microwave power, solvent concentration, extraction time, and their interactions and was estimated by least square regression. From the literature it was found that Hong et al. (2001) used microwave power of 150–300 W and time from 20 to 200 s to extract phenolic compounds from grape seed. Mayer et al. (2008) used microwave power of 150 W and 200 s for extraction of proanthocyanidins from grape seed. Thus 100–300 W and 1–5 min were taken as the minimum and maximum values for the screening design. A response surface study was initially conducted using a two-level, three-factorial design for screening the effects of the factors, ethanol concentration of 50% and 90%, microwave power levels of 100 and 300 W, and extraction time of 1 and 5 min on the extraction yields of total phenols, the antioxidant activity, and the color of the extracts. Then the design was augmented to central composite design by the addition of center and axial points.

## Materials and Methods

### Materials and Reagents

*V. vinifera* (var. Bangalore blue), a variety of grapes, were procured from the local markets of Coimbatore, Tamil Nadu, India. This variety of grapes is widely grown in the southern states of India like Karnataka and Tamil Nadu. The grapes were processed for the production of juice in the Post Harvest Technology Centre, Tamil Nadu Agricultural University, Coimbatore. The residual pomace was collected and the seeds were manually removed from the pomace. The seeds were cleaned, lyophilized in a batch-type freeze dryer (Ilshin, version 1.0, Model FD5505) until constant weight was obtained. They were packed in polyethylene bags and stored at  $-18^{\circ}\text{C}$  until further analysis. All chemicals used such as Folin–Ciocalteu reagent, gallic acid, catechin and tannic acid, sodium carbonate, 1,1-diphenyl-2-picrylhydrazyl (DPPH), and 2,4,6-tripyridyl-*s*-triazine were of analytical and HPLC grade obtained from Sigma Chemicals Co (St. Louis, MO, US).

### Sample Preparation

Dried grape seeds were milled in a coffee grinder (Proctor Silex, Model E167C) for 2 min, the process was stopped at

15-s interval for 15 s to avoid heating of the sample (Gokturk Baydar et al. 2007). The grape seed powder (10 g) was defatted with 30 ml of hexane in an Erlenmeyer flask at  $60^{\circ}\text{C}$  for 6 h. The hexane containing the lipid portion was filtered, and the grape seed powder was kept in the fume hood to allow the residual hexane to evaporate.

### Microwave Extraction of Phenolic Antioxidants

Dry defatted grape seed powder ( $0.5\pm 0.05$  g) was placed in a 250-ml quartz vessel. Fifty milliliters of different ethanol concentrations (30%, 45%, 60%; v/v) was used for extraction. The vessel was introduced into the microwave cavity and fitted with a condenser. A focused-type, open-vessel microwave system (Star System 2, CEM Matthews, USA) operating at 800 W maximum power with a frequency of 2,450 MHz was used. The microwave power applied was intermittent with power on for 30 s/min. The extracts were centrifuged for 20 min and filtered with 0.22- $\mu\text{m}$  filters (Whatman Puradisc 13-mm nylon syringe filters) for the determination of total phenols, color, and antioxidant activity.

### Experimental Design

Face-centered central composite design (CCD) was used to determine the optimal conditions of microwave-assisted extraction of total phenols (TP), antioxidants activity (ferric ion reducing antioxidant power (FRAP) and DPPH), and color from grape seeds. Central composite design is highly efficient and provides sufficient information on the effect of process variables for resourceful optimization with reduced number of total experimental runs. The three independent factors studied were the microwave power,  $X_1$  (100, 150, and 200 W), extraction time,  $X_2$  (2, 4, and 6 min), and ethanol concentration,  $X_3$  (30%, 45%, and 60% ethanol), which were coded at three levels as shown in Table 1.

The CCD included eight factorial points, six axial points, and three center points, which totals 17 experimental runs, and was used to fit the second-order polynomial model

**Table 1** Independent variables in CCRD for optimization of grape seed extract

Independent variables	Coded levels			
	-1	0	1	
Microwave power (W)	$X_1$	100	150	200
Extraction time (min)	$X_2$	2	4	6
Ethanol concentration (% v/v)	$X_3$	30	45	60

Power (100 W=14%, 150 W=20%, 200 W=27% of maximum power)

similar to that shown in Eq. 1 (Haaland 1989, Abdeshahian et al. 2010).

$$Y = \beta_0 + \sum_{i=1}^3 \beta_i x_i + \sum_{i=1}^3 \beta_{ii} x_i^2 + \sum_{i < j=1}^3 \sum \beta_{ij} x_i x_j \quad (1)$$

where  $Y$  is the predicted response,  $\beta_0$  is the regression coefficient for the intercept (a constant),  $\beta_i$  is the coefficient for the linear effect,  $\beta_{ii}$  is the coefficient for the quadratic effect,  $\beta_{ij}$  is the coefficient for the interaction effect, of variables  $i$  and  $j$ ,  $X_i$ , and  $X_j$  are independent variables. The JMP software Version 8 (SAS Institute Inc., Cary, NC, USA) was used for the statistical analysis and to generate the response surface plots for the quadratic models.

### Analysis of Total Phenolic Compounds

A colorimetric method using Folin–Ciocalteu assay was used to determine the total phenols content in the grape seed extract. The method used by Jayaprakasha et al. (2003), Jayaprakasha and Jaganmohan Rao (2000), and Singh et al. (2011) was used with slight modifications. One milliliter of the grape seed extract having 1 mg of grape seed/ml of ethanol was mixed with 7.5 ml of double deionized water, followed by the addition of 0.5 ml of Folin–Ciocalteu reagent, and 1 ml of 7.5% sodium carbonate solution. The mixture was incubated at room temperature in the dark for the development of color for 30 min. The absorbance was measured at 765 nm using a spectrophotometer (Ultrospec 2100 pro, Biochrom Ltd., Cambridge, England). The total phenols concentration was expressed in terms of gallic acid equivalents (GAE), catechin equivalents (CAT), and tannic acid equivalents (TAE) based on the standard curves given in Table 2 which were plotted for 30%, 45%, and 60% ethanol concentrations in respective equivalents ( $Y$ =concentration,  $x$ =absorbance at 765 nm).

### Determination of Antioxidant Activity

Several methods are used to evaluate antioxidant activities of natural compounds in foods or biological systems with varying results. Depending upon the reactions involved, these assays can roughly be classified into two types: assays based on hydrogen atom transfer reactions and assays based on electron transfer (ET). The ET-based assays measure the capacity of an antioxidant in the reduction of an oxidant, which changes color when reduced. The degree of color change is correlated with the sample's antioxidant concentrations. The ET-based assays include FRAP and DPPH. The antioxidant activity of the microwave-assisted extracts of grape seed was determined by both FRAP and DPPH.

### Antiradical Scavenging Activity

The DPPH assay was used to determine the free radical scavenging activity of grape seed extract. The method by Nair et al. (2007) was adopted with some modifications (Singh et al. 2011). An aliquot of 50  $\mu$ l ethanolic extract of grape seed having 1 mg of GS/ml of ethanol was added to 1.5 ml of DPPH (3.94 mg/100 ml ethanol). The discoloration of purple is due to the paring of the free electron with the free radical scavenger present in DPPH. The color change was measured with a spectrophotometer (Ultrospec 2100 pro, Biochrom Ltd., Cambridge, England) at 517-nm absorbance after 20 min. The percentage of free radicals scavenged by DPPH radical was calculated using the following equation:

$$\text{DPPH(\% inhibition)} = \left[ \frac{(\text{Abs control (517nm)} - \text{Abs sample (517nm)})}{\text{Abs control (517nm)}} \right] \times 100 \quad (2)$$

where, Abs control = absorption of blank sample and Abs sample = absorption of grape seed extract.

**Table 2** Standard curve for different solvent concentrations in terms of standard equivalents

Sample no.	Standard curve	Equation	$R^2$
1	Gallic acid standard curve, 30% EtOH	$Y = 81.75x - 0.040$	0.997
2	Gallic acid standard curve, 45% EtOH	$Y = 91.13x - 0.052$	0.999
3	Gallic acid standard curve, 60% EtOH	$Y = 92.58x - 0.019$	0.999
4	Catechin standard curve, 30% EtOH	$Y = 50.39x + 0.036$	0.999
5	Catechin standard curve, 45% EtOH	$Y = 51.57x + 0.045$	0.991
6	Catechin standard curve, 60% EtOH	$Y = 49.17x + 0.091$	0.991
7	Tannic acid standard curve, 30% EtOH	$Y = 69.18x + 0.052$	0.997
8	Tannic acid standard curve, 45% EtOH	$Y = 69.87x + 0.018$	0.999
9	Tannic acid standard curve, 30% EtOH	$Y = 87.66x + 0.002$	0.993

$x$  concentration (in milligrams/milliliters),  $Y$  absorbance at 765 nm



## Reducing Power

The FRAP assay takes advantage of electron transfer reactions and is used to measure the total antioxidant activity of natural compounds. Herein a ferric salt, Fe(III)(TPTZ)<sub>2</sub>Cl<sub>3</sub> (TPTZ) 2,4,6-tripyridyl-*s*-triazine is used as an oxidant. Antioxidants are used as reductants in a redox-linked colorimetric method. The method used by Benzie and Strain (1996) and Deighton et al. (2000) was used in this experiment with some modifications. Acetic acid buffer of 250 ml, 25 ml of TPTZ (2,4,6-tripyridyl-*s*-triazine) and 25 ml of ferric chloride solution was added together at 10:1:1 ratio, to make the FRAP reagent. Grape seed ethanolic extract (50 μl) was added to 1.5 ml of FRAP reagent, vortexed for 60 s, and allowed to stand for 6 min at room temperature. The control consisted of 1.5 ml FRAP reagent to which were added 50 μl distilled water. The color change was read with a spectrophotometer at 593 nm. Ascorbic acid, which is an antioxidant, was used as a standard solution to determine the reactivity of FRAP solutions using the standard curve shown in Eq. 3 ( $R^2=0.997$ ).

Absorbance of grape seed extract (593nm)

$$= 244.7 \times \text{Conc. of grape seed extract} - 0.014 \quad (3)$$

## Determination of Extract Color

The color of the extract was measured using a colorimeter (Konica Minolta, Japan), equipped with 1 cm measuring head. The colorimeter works on the principle of focussing the light and measuring the energy reflected from the sample across the entire visible spectrum. The color meter relies on “standard observer curves” that define the quantity of red, green, and blue colors. The primary light requires matching a series of colors across the visible spectrum and the mathematical model used to describe these colors as  $L^*a^*b^*$  color space. The colorimeter provides reading in terms of  $L^*$ ,  $a^*$ , and  $b^*$ , where luminance ( $L$ ) forms the vertical axis, which indicates whiteness to darkness. The chromatic portion of the solids is defined by:  $a$  (+) redness,  $a$  (−) greenness,  $b$  (+) yellowness, and  $b$  (−) blueness.

The colorimeter was calibrated using the standard white calibration plate provided by the manufacturer before the start of the measurements. Buci Koji et al. (2009) and Wrolstad et al. (2005) similarly used the  $L$ ,  $a$ , and  $b$  system to determine the color changes of grape seed extract and anthocyanin pigments. One milliliter of the microwave-assisted grape seed extracts was placed in small vials of 1 cm in diameter to match the orifice of the colorimeter. This setup was covered with a black cover in order to prevent the scattering of light.  $L^*$  (light/dark),  $a^*$  ( $a^+$ , redness/green −  $a^-$ ) and  $b^*$  ( $b^+$ , yellowness/blueness −  $b^-$ ) values were recorded

for each sample. From the obtained  $L^*$ ,  $a^*$ , and  $b^*$  values, the total change in color of the extract ( $\Delta E$ ) was calculated using the following equation:

$$\Delta E = \sqrt{(L_0 - L^*)^2 + (a_0 - a^*)^2 + (b_0 - b^*)^2} \quad (4)$$

Where  $L_0$ ,  $a_0$ , and  $b_0$  denotes the value of pure ethanol at 30%, 45%, and 60%, v/v used as standard reference to compare the color change of the ethanolic extract.

## Statistical Analysis

Statistical analysis was performed using the JMP 8 software (SAS Institute Inc., Cary, NC, USA). The data were analyzed by analysis of variance (ANOVA) and the adequacy of the response surface model was determined by evaluating the lack of fit and coefficient of determination ( $R^2$ ). The statistical significance of the model and its variables was determined at 5% probability level ( $p < 0.05$ ). The optimal conditions for MAE of grape seed phenolic antioxidants were obtained based on modeling and desirability function that could be visually explained in terms of three-dimensional response surface plots and contour plots.

## Results and Discussion

### Fitting the Model

A two-level, three-factorial design was used for screening the factors. The studied factors and their levels were: the microwave power at 100 and 300 W, the extraction time of 1 and 5 min, and the concentration of ethanol 50% and 90%, v/v. The results obtained indicated that the yield of total phenolics can be maximized by increasing the extraction time and decreasing the solvent concentration. At increased microwave power level of 300 W in the screening experiment, rapid heating of the extraction medium was seen. Due to high cavitation, bubbling of the substance occurred. This led to the entry of the extraction medium into the condenser. Based on the observation from the screening design, the microwave level was fixed between 100 and 200 W in the CCD. Hence, for further experiments, it was decided to modify the experimental domain as follows: microwave powers between 100 and 200 W, extraction time of 2 and 6 min, and concentration of ethanol of 30% and 60%. The different combinations of the independent variables and their corresponding response in terms of total phenols are presented in Table 3.

### Total Phenols

Response surface regression models were fitted to the experimental data. The ANOVA of the quadratic regression



**Table 3** Face-centered central composite design (CCD) with observed response for total phenols from microwave-assisted extraction of grape seeds

Design points	MWP, Watt	Time, Min	Sol. conc. % v/v	Total phenols (mg/g)								
				GAE			CAT			TAE		
				Actual	Predicted	STD error	Actual	Predicted	STD error	Actual	Predicted	STD error
R1	200	2	30	10.373	10.470	±0.165	16.829	16.977	±0.292	12.258	12.324	±0.230
R2	100	6	30	12.269	12.345	±0.165	19.905	20.034	±0.292	14.498	14.580	±0.230
R3	200	6	30	13.235	13.289	±0.165	21.473	21.592	±0.292	15.640	15.777	±0.230
R4	100	2	30	11.351	6.423	±1.238	18.416	9.760	±2.182	13.414	7.067	±1.720
R5	100	2	60	9.138	9.084	±0.165	17.206	17.087	±0.292	9.651	9.514	±0.230
R6	200	2	60	9.689	9.613	±0.165	18.243	18.114	±0.292	10.233	10.151	±0.230
R7	100	6	60	10.510	10.413	±0.165	19.788	19.640	±0.292	11.100	11.034	±0.230
R8	200	6	60	15.241	7.839	±1.238	28.696	15.008	±2.182	16.096	7.610	±1.720
R9	100	4	45	9.920	9.996	±0.145	17.530	17.669	±0.255	12.938	13.059	±0.201
R10	200	4	45	10.809	10.733	±0.145	19.100	18.961	±0.255	14.098	13.977	±0.201
R11	150	2	45	9.580	9.613	±0.145	16.928	17.027	±0.255	12.495	12.647	±0.201
R12	150	6	45	11.720	11.687	±0.145	20.710	20.611	±0.255	15.286	15.134	±0.201
R13	150	4	30	11.755	11.527	±0.145	19.071	18.675	±0.255	13.891	13.606	±0.201
R14	150	4	60	9.905	10.133	±0.145	18.650	19.046	±0.255	10.461	10.746	±0.201
R15	150	4	45	10.995	10.955	±0.104	19.430	19.359	±0.183	14.341	14.289	±0.144
R16	150	4	45	10.886	10.955	±0.104	19.236	19.359	±0.183	14.198	14.289	±0.144
R17	150	4	45	10.984	10.955	±0.104	19.411	19.359	±0.183	14.327	14.289	±0.144

GAE gallic acid equivalents, CAT catechin equivalents, TAE tannic acid equivalents

model for total phenols expressed in gallic acid equivalents is given in Table 4. The model obtained for total phenols in terms of gallic acid equivalents was significant with  $p$  value=0.0002, with  $R^2$  value of 0.9906. It was found that there

**Table 4** ANOVA for total phenols by MAE of grape seeds

Total phenols (mg gallic acid equivalents g <sup>-1</sup> dw)					
Source	SS	df	MS	F value	p value
Model	17.087250	9	1.89858	58.9808	0.0002 <sup>a</sup>
Lack of fit	0.15374735	3	0.051249	14.2319	0.0664
Error	0.160949	5	0.03219		
Term	Estimate	Std error	t ratio	F value	p value
Intercept	10.955	0.103586	105.76		<0.0001 <sup>a</sup>
MWP (100, 200)	0.36855	0.069487	5.30	28.1308	0.0032 <sup>a</sup>
Time (2, 6)	1.03705	0.069487	14.92	222.7352	<0.0001 <sup>a</sup>
Sol conc. (30, 60)	-0.69735	0.069487	-10.04	100.7143	0.0002 <sup>a</sup>
MWP × time	-0.77575	0.315043	-2.46	6.0632	0.0571
MWP × sol conc.	-0.8795	0.315043	-2.79	7.7935	0.0384 <sup>a</sup>
Time × sol conc.	-1.14825	0.315043	-3.64	13.2841	0.0148 <sup>a</sup>
MWP × MWP	-0.5905	0.163783	-3.61	12.9988	0.0155 <sup>a</sup>
Time × time	-0.305	0.163783	-1.86	3.4679	0.1216
Sol conc. × sol conc.	-0.125	0.163783	-0.76	0.5825	0.4798
R <sup>2</sup>	0.990669				
Adjusted R <sup>2</sup>	0.973872				

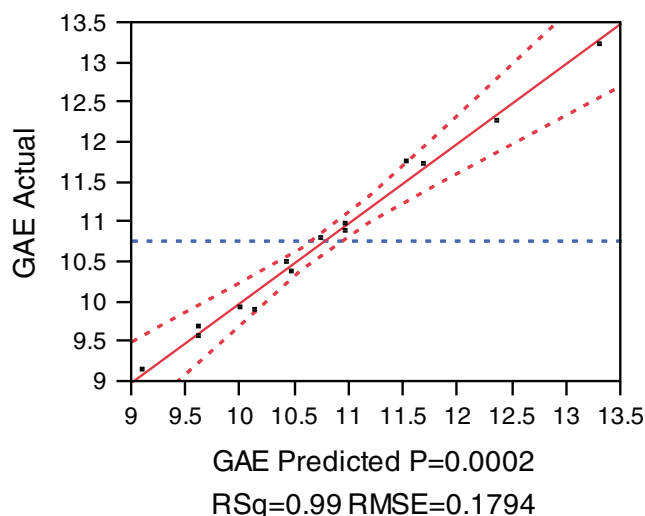
<sup>a</sup>Significant

is a reasonable agreement with the adjusted  $R^2$  of 0.9738 in the model. The lack of fit was not significant ( $p > 0.05$ ), suggesting that the model was well fitted and could be used to predict the total phenols from microwave-assisted grape seed extract.

The model indicated that the linear effects of extraction time and solvent concentration had the greatest significance on total phenols of the extract. Values of ( $p$ ) less than 0.05 indicate the significant terms in the model, based on which the linear terms such as microwave power, time of extraction, solvent concentration, bilinear terms such as  $MWP \times Sol. Conc.$ ,  $Time \times Sol. Conc.$ , and quadratic terms of  $MWP^2$ , were found to be significant. The predictive, second-order, polynomial model for total phenols content of microwave-assisted grape seed extracts, in terms of coded factors' levels, is shown in Eq. 5.

$$\begin{aligned}
 TP(\text{GAEmg/g dw GS}) = & +10.955 + 0.368(\text{MWP}) \\
 & + 1.03705(\text{Time}) \\
 & - 0.697(\text{Sol.Conc.}) \\
 & - 0.775(\text{MWP} \times \text{Time}) \\
 & - 0.879(\text{MWP} \times \text{Sol.Conc.}) \\
 & - 1.148(\text{Time} \times \text{Sol.Conc.}) \\
 & - 0.590(\text{MWP}^2) \\
 & - 0.305(\text{Time}^2) \\
 & - 0.125(\text{Sol.Conc.}^2) \quad (5)
 \end{aligned}$$

The Pearson's correlation coefficient for actual vs. predicted value is  $r=0.995$  (Fig. 2). Hence this model can be used to predict total phenolics under different experimental conditions during microwave-assisted extraction.



**Fig. 2** Predicted (in milligrams GAE/g) vs. actual (in milligrams GAE/g) for total phenols in MAE of grape seed

Similarly, a regression analysis was carried out to quantify the total phenols in terms of catechin equivalents. The response predictions of total phenols in terms of catechin equivalents was significant in the given model with  $p=0.0009$  (Eq. 6). The  $R^2$  value for the model was 0.9806 with a reasonable agreement with the adjusted  $R^2$  of 0.9457, and the lack of fit was not significant 0.0678 ( $p > 0.05$ ) suggesting that the model could be used to predict the response. It was found that microwave power, time and the interaction of MWP and solvent concentration, time and solvent concentration, and  $MWP^2$  were significant factors in the model. The predicted response for total phenols in terms of catechin equivalents, in term of coded factors level, is given in the Eq. 6 neglecting some insignificant terms:

$$\begin{aligned}
 TP(\text{CATmg/g dw GS}) = & +19.359 + 0.646(\text{MWP}) \\
 & + 1.791(\text{Time}) \\
 & + 0.185(\text{Sol.Conc.}) \\
 & - 1.414(\text{MWP} \times \text{Time}) \\
 & - 1.547(\text{MWP} \times \text{Sol.Conc.}) \\
 & - 1.930(\text{Time} \times \text{Sol.Conc.}) \\
 & - 1.044(\text{MWP}^2) \\
 & - 0.54(\text{Time}^2) \\
 & - 0.498(\text{Sol.Conc.}^2) \quad (6)
 \end{aligned}$$

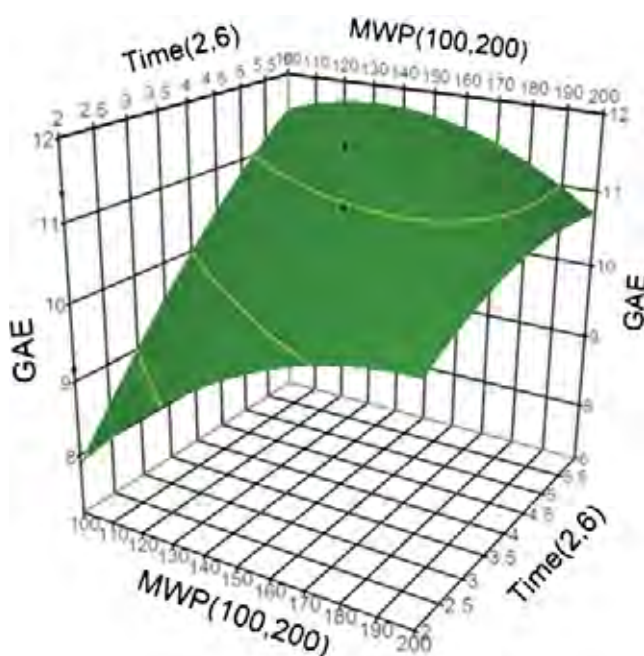
The predictive response of total phenols in terms of tannic acid equivalents by regression analysis was found to be significant ( $p$  value  $< 0.0001$ ). The  $R^2$  value of the model was found to be 0.9939 with a reasonable agreement with the adjusted  $R^2$  of 0.9829 and insignificant lack of fit ( $p > 0.05$ ). The final equation in terms of coded factors' levels for total phenols is:

$$\begin{aligned}
 TP(\text{TAEmg/g dw GS}) = & +14.288 + 0.458(\text{MWP}) \\
 & + 1.243(\text{Time}) \\
 & - 1.429(\text{Sol.Conc.}) \\
 & - 1.015(\text{MWP} \times \text{Time}) \\
 & - 1.155(\text{MWP} \times \text{Sol.Conc.}) \\
 & - 1.498(\text{Time} \times \text{Sol.Conc.}) \\
 & - 0.770(\text{MWP}^2) \\
 & - 0.398(\text{Time}^2) \\
 & - 2.112(\text{Sol.Conc.}^2) \quad (7)
 \end{aligned}$$

The predicted vs. actual plot obtained for total phenols in terms of catechin equivalents and tannic acid equivalents were similar to that obtained for gallic acid equivalents (Fig. 2) with Pearson's correlation coefficient ( $r$ ) for GAE  $r=0.995$ , CAT  $r=0.990$ , TAE  $r=0.996$ . Thus the given models can be used to predict the total phenols of microwave-assisted extraction of grape seeds in milligrams equivalents per gram of dry weight (dw) grape seed, respectively.

#### Effect of Process Variables on Total Phenols

The relationship between independent and dependent variable can be well illustrated with three-dimensional response surfaces. Figure 3 gives the interaction of MWP and time at a fixed solvent concentration of 45% ethanol concentration. It was found that there was an increase in total phenols with an increase in time of extraction. By keeping time constant with constant solvent concentration and increasing the MWP, there was a gradual increase in the total phenols. The maximum total phenol from the extract was obtained with increased extraction time. It revealed that by decreasing the solvent concentration from 60% to 30% and by increasing the MWP from 100 to 170 W, it increased the total phenols. Wang et al. (2010) reported that 70% ethanol concentration, microwave power 255 W, and extraction time 6.5 min gave maximum flavonoids from *Radix puerariae*. MAE of secoisolariciresinol diglucoside from flaxseed by Nemes and Orsat (2009) reported MWP of 135 W, sodium hydroxide concentration of 0.5 M, and 3 min extraction time which gave 10% increase in the extraction yield when



**Fig. 3** Response surface plot for total phenols in terms of GAE

compared to the microwaveless control method. Pan et al. (2000) studied the MAE of glycyrrhizic acid from licorice root with extraction times of 4–5 min, ethanol concentrations of 50–60% (v/v).

Similar surface plots were obtained for total phenols in terms of tannic acid equivalents and catechin equivalents. The optimum extraction conditions for maximum total phenols from grape seed was found at 170 W, 6 min, and 30% ethanol concentration yielding  $13.5 \pm 0.48$  mg GAE/g dw GS with a desirability of 0.982. The microwave power of 168 W, 6 min, and 30% ethanol concentration yielded more total phenols in terms of catechin equivalents at  $22 \pm 0.86$  mg CAT/g dw GS with the desirability of 0.997. The predicted optimum extraction conditions in terms of tannic acid equivalents were 160 W, 6 min, and 33.7% ethanol concentration with 0.992 desirability giving  $16.1 \pm 0.56$  mg TAE/g dw GS.

Ghafoor et al. (2009) studied the extraction of phenols and antioxidant from grape seed by ultrasound-assisted extraction. The surface plots obtained were similar to the present study with 53.15% ethanol concentration v/v, 56.03 °C, and 29.03 min of extraction time yielding maximum total phenols of 5.44 mg GAE/100 ml. In the present study with microwave-assisted extraction, it was possible to obtain a high amount of total phenols in less time (6 min) with reduced 30–34% (v/v) of ethanol. The amount of total phenols obtained by Li et al. (2008) was in the range of 7.5 to 44 mg GAE/100 g dry matter. Mandic et al. (2008) reported 3.59 to 11.7 g GAE/kg of dry grape seed using ethanolic solvent for extraction. Jayaprakasha et al. (2001) reported a range of 18.2 to 20.8 CAT/100 g grape seed extract for different ratios of ethanol water mixture. Since different grape variety, solvents combinations, and methods of extraction were used, the comparison of the results can only be informative. The variation in the results depends upon factors such as concentration of phenolic compounds in different grape cultivars, and it is also influenced by viticulture practices and environmental factors such as maturity stage, area of production, and seasonal conditions (Gómez-Alonso et al. (2007), Cheynier and Rigaud (1986), Mazza and Francis (1995), Koundouras et al. (2006)).

#### Antioxidant Activity

The antioxidant activity in terms of DPPH and FRAP for different combinations of the independent variables are presented in Table 5.

#### Antiradical Activity

The radical scavenging activity of the grape seed extract was determined by the DPPH method. To increase the fitness of the model, log transformation of the response was done. Transformation of the variables is said to enhance the

**Table 5** Observed responses for the antioxidant activity and color value of the microwave-assisted extraction of grape seed in the experimental design

	Antioxidant activity						Color value			
	DPPH			FRAP			$L^*$	$a^*$	$b^*$	$\Delta E$
	Actual	Predicted	STD error	Actual	Predicted	STD error				
R1	67.113	67.173	±0.012	115.803	115.883	±0.010	31.110	-1.327	2.167	2.421
R2	73.065	85.081	±0.025	124.210	141.193	±0.021	29.630	-1.087	2.940	3.996
R3	77.232	76.922	±0.011	130.095	129.644	±0.010	30.350	-0.940	2.663	3.233
R4	56.994	56.736	±0.011	101.512	101.091	±0.010	35.417	-0.813	1.453	2.555
R5	79.464	80.007	±0.012	133.247	134.008	±0.010	30.840	-0.393	3.860	3.352
R6	80.357	80.581	±0.012	134.508	134.807	±0.010	27.977	-0.643	4.640	5.821
R7	79.762	79.913	±0.011	133.668	133.872	±0.010	28.180	-0.530	4.860	5.813
R8	61.012	61.461	±0.012	107.186	107.871	±0.010	28.217	-0.437	3.683	5.075
R9	79.464	79.135	±0.010	133.247	132.839	±0.008	32.387	-0.433	3.083	2.627
R10	76.042	75.514	±0.009	128.414	127.669	±0.008	32.187	-0.863	3.437	3.075
R11	73.661	73.225	±0.009	125.051	124.495	±0.008	32.690	-0.393	3.233	2.715
R12	78.720	78.313	±0.010	132.197	131.613	±0.008	33.600	-0.400	2.377	1.924
R13	71.131	71.680	±0.010	121.478	122.323	±0.008	34.240	-0.633	2.520	2.155
R14	77.530	76.086	±0.009	130.515	128.467	±0.008	28.350	-0.573	4.550	5.479
R15	76.190	77.056	±0.005	128.624	129.810	±0.005	28.217	-0.350	4.047	5.956
R16	77.083	77.056	±0.005	129.885	129.810	±0.005	28.113	-0.376	4.060	6.048
R17	76.190	77.056	±0.005	128.624	129.810	±0.005	28.583	-0.420	4.457	5.928

$\Delta E$  total change in color of the extract, DPPH (percent inhibition), FRAP (in micromoles of ascorbic acid eq/g dw)

interpretability of the results (Zumbo and Harwell 1999). A regression model was used to fit the DPPH data, and the ANOVA results are presented in Table 6. It was found that linear terms such as microwave power, time of extraction, and solvent concentration and quadratic terms indicating interactions such as MWP × time, MWP × Sol. Conc., Time × Sol. Conc., and Sol. Conc.<sup>2</sup> were highly significant factors ( $p < 0.05$ ). The model was significant with  $p$  value  $< 0.0001$  and a nonsignificant lack of fit 0.1825 ( $p > 0.05$ ). The  $R^2$  value of the model was 0.9937 and the adjusted  $R^2$  was 0.9843. Thus this model can be used to predict the radical scavenging activity for the microwave-assisted extraction of grape seed extract. The predictive model in terms of coded factors' levels for DPPH is given in Eq. 8.

Log DPPH (% inhibition)

$$\begin{aligned}
 &= +4.3445 - 0.0234(\text{MWP}) + 0.0335(\text{Time}) \\
 &+ 0.0298(\text{Sol. Conc.}) - 0.0674(\text{MWP} \times \text{Time}) \\
 &- 0.0404(\text{MWP} \times \text{Sol. Conc.}) \\
 &- 0.1015(\text{Time} \times \text{Sol. Conc.}) + 0.0032(\text{MWP}^2) \\
 &- 0.0174(\text{Time}^2) - 0.0424(\text{Sol. Conc.}^2) \quad (8)
 \end{aligned}$$

Reducing Power

FRAP assay was used to determine the reducing power of the antioxidants present in microwave-assisted grape seed extracts. Similar to DPPH, the best fit to the FRAP data was obtained with log transformation. The model  $p$  value  $< 0.0001$  implies the model is significant. The lack of fit was not significant, 0.1831 ( $p > 0.05$ ). The  $R^2$  value of the model is 0.9935 and the adjusted  $R^2$  of 0.9838. It was found that the linear and the quadratic terms which were significant in DPPH model were similar for the FRAP model. The values obtained for antioxidant activity from DPPH and FRAP showed significant correlation ( $r = 1$ ). The predicted response for FRAP in terms of coded values for factors' levels is given in Eq. 9.

Log FRAP(( $\mu$  M of Ascorbic acid eq)/g dw)

$$\begin{aligned}
 &= +4.8660 - 0.0198(\text{MWP}) + 0.0277(\text{Time}) \\
 &+ 0.0245(\text{Sol. Conc.}) - 0.0554(\text{MWP} \times \text{Time}) \\
 &- 0.0326(\text{MWP} \times \text{Sol. Conc.}) - 0.0837(\text{Time} \\
 &\times \text{Sol. Conc.}) + 0.0032(\text{MWP}^2) \\
 &- 0.0140(\text{Time}^2) - 0.0349(\text{Sol. Conc.}^2) \quad (9)
 \end{aligned}$$

**Table 6** ANOVA for antioxidant activity of MAE of grape seed

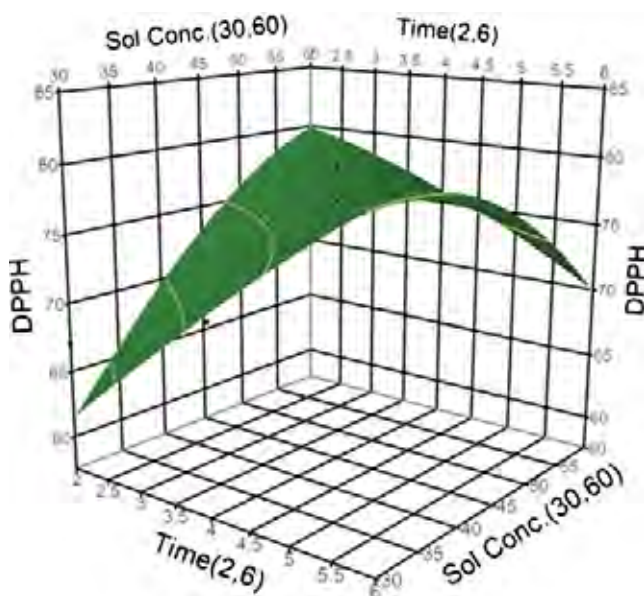
		Antioxidant activity									
		DPPH (% inhibition)					FRAP ( $\mu\text{M}$ of ascorbic acid eq)/g dw				
Source	SS	df	MS	F value	p value	SS	df	MS	F value	p value	
Model	0.15008640	9	0.016676	105.9275	<0.0001 <sup>a</sup>	0.10116425	9	0.011240	102.2450	<0.0001 <sup>a</sup>	
Lack-of-fit	0.00085406	4	0.000214	4.7175	0.1825	0.00059617	4	0.000149	4.6977	0.1831	
Error	0.00094459	6	0.000157			0.00065962	6	0.000110			
Term	Estimate	Std Error	t ratio	F value	p value	Estimate	Std Error	t ratio	F value	p value	
Intercept	4.3445263	0.005425	800.79		<0.0001 <sup>a</sup>	4.8660683	0.004534	1073.3		<0.0001 <sup>a</sup>	
MWP (100, 200)	-0.023418	0.004839	-4.84	23.4238	0.0029 <sup>a</sup>	-0.019847	0.004043	-4.91	24.0930	0.0027 <sup>a</sup>	
Time (2, 6)	0.0335854	0.004839	6.94	48.1792	0.0004 <sup>a</sup>	0.0277993	0.004043	6.88	47.2688	0.0005 <sup>a</sup>	
Sol conc (30, 60)	0.02983	0.004839	6.16	38.0072	0.0008 <sup>a</sup>	0.0245048	0.004043	6.06	36.7291	0.0009 <sup>a</sup>	
MWP $\times$ time	-0.067422	0.005627	-11.98	143.5712	<0.0001 <sup>a</sup>	-0.055474	0.004702	-11.80	139.1828	<0.0001 <sup>a</sup>	
MWP $\times$ sol conc.	-0.04043	0.005627	-7.19	51.6272	0.0004 <sup>a</sup>	-0.032653	0.004702	-6.94	48.2241	0.0004 <sup>a</sup>	
Time $\times$ sol conc.	-0.101595	0.005627	-18.06	325.9954	<0.0001 <sup>a</sup>	-0.083779	0.004702	-17.82	317.4527	<0.0001 <sup>a</sup>	
MWP $\times$ MWP	0.0032098	0.007786	0.41	0.1699	0.6945	0.0032216	0.006506	0.50	0.2452	0.6381	
Time $\times$ time	-0.0174	0.007786	-2.23	4.9940	0.0668	-0.014003	0.006506	-2.15	4.6318	0.0749	
Sol conc. $\times$ sol conc.	-0.042491	0.007786	-5.46	29.7822	0.0016 <sup>a</sup>	-0.0349	0.006506	-5.36	28.7710	0.0017 <sup>a</sup>	
R <sup>2</sup>	0.993746					0.993522					
Adjusted R <sup>2</sup>	0.984364					0.983805					

<sup>a</sup> Significant

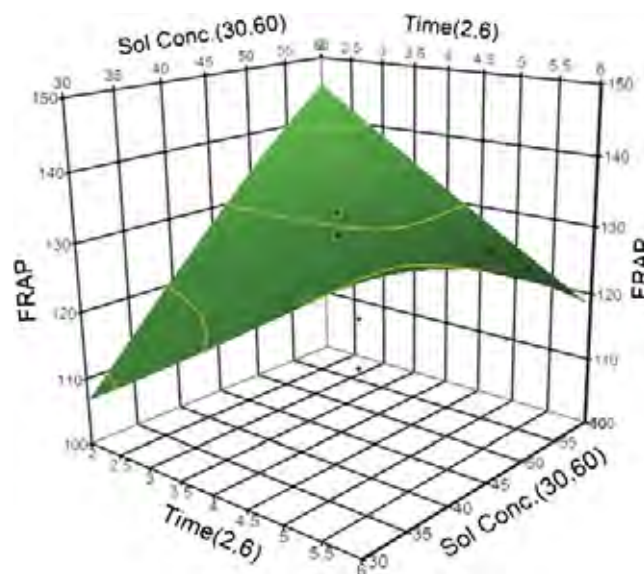


## Effect of Process Variables on Antioxidant Activity

It was found that by increasing the time of extraction with decrease in solvent concentration and MWP, the antioxidant activity of the extract increased in terms of DPPH and FRAP. Figures 4 and 5 represent the response surface plot for antioxidant activity in terms of DPPH percent inhibition and FRAP. The Pearson's correlation coefficient for DPPH and FRAP was 0.996. It was found that microwave power, time of extraction and solvent concentration and the interaction of microwave power with time and solvent concentration, and interactions of time and solvent concentration played a significant role in the antioxidant activity of DPPH and FRAP. The diffusion of the antioxidants from the solids matrix to the solvent is enhanced when the time of extraction increases. Grape seeds are natural reservoir of low molecular weight catechins (monomers, dimmers, trimers, and oligomers) and flavonoids with orthodiphenolic structure in the B ring (Katalinic et al. 2004). Spranger et al. (2008) reported that the radical scavenging activity of grape seed by DPPH is positively related to their degree of polymerization, i.e., polymer > oligomer > monomer (catechin). As grape seed has high proanthocyanidin content, it has higher antioxidant activity when compared to vitamin C and vitamin E (Spranger et al. 2008, Ramchandani et al. 2010). Since the structural features of phenolic compounds are reported responsible for the antioxidant activity, measurement of total phenols may be related to their antioxidant activity (Katalinic et al. 2004). The methods for assessing antioxidant activity vary considerably which depends upon the type of radical generated from a given compound; hence,

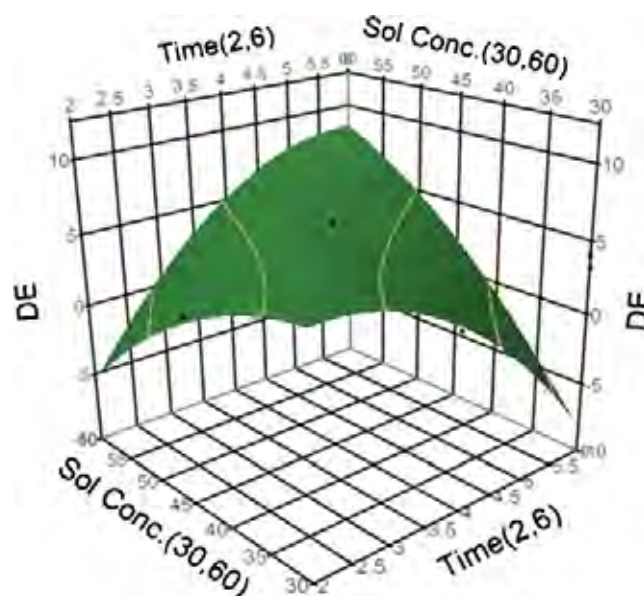


**Fig. 4** Response surface plot for antioxidant activity in DPPH percent inhibition



**Fig. 5** Response surface plot for antioxidant activity in FRAP

the results obtained by different methods are not always comparable (Spranger et al. 2008). But similar results were obtained by Ramchandani et al. (2010) for grape seed (Bangalore blue variety) with a percent inhibition range from 50% to 80%, proving that this variety of grape seed has a high potential antioxidant activity and could efficiently be extracted by microwave-assisted extraction. Pan et al. (2008) studied the antioxidant activity of microwave-assisted extract of longan peel and found that antioxidant of microwave-assisted extract of Longan peel was superior to Soxhlet extract of Longan peel. This suggests that microwave-assisted



**Fig. 6** Response surface plot for total change in color of extract ( $\Delta E$  or DE)

extraction of antioxidants could be an effective method with respect to time and less solvent wastage. The highly significant correlation coefficient was obtained for FRAP and DPPH data ( $r=1$ ). This indicates that both FRAP and DPPH can be successfully used to obtain the antioxidant activity of the grape seed extract.

Effect of Process Variables on Color

Color denotes visual appearance of the product, whereas pigments or colorants are the chemical compounds that impart the observed color (Wrolstad et al. 2005). The color measurements are presented in Table 5. It was found that solvent concentration played a significant role in the color of the extract as shown in Fig. 6 which represents the response surface graph for total change in color. It was found that

increase in the solvent concentration and extraction time there was an increase in the color of the extract. Canals et al. (2005) stated that when the ethanol concentration is high, the extraction of proanthocyanidins from skin and seed significantly increases. The color of the extract with 45% and 60% ethanol concentration had more intense color than 30% EtOH concentration. The values obtained in this study were similar to the values obtained by Buci Koji et al. (2009) who studied the influence of solvent and temperature on extraction of phenolic compounds and color of the extracts from grape seed.

The color of the extract is directly related to the presence of anthocyanin color pigments and may be due to the interactions of non-pigmented compounds like catechin and epicatechin (Buci Koji et al. 2009). Stintzing et al. (2002) studied the color and antioxidant properties of

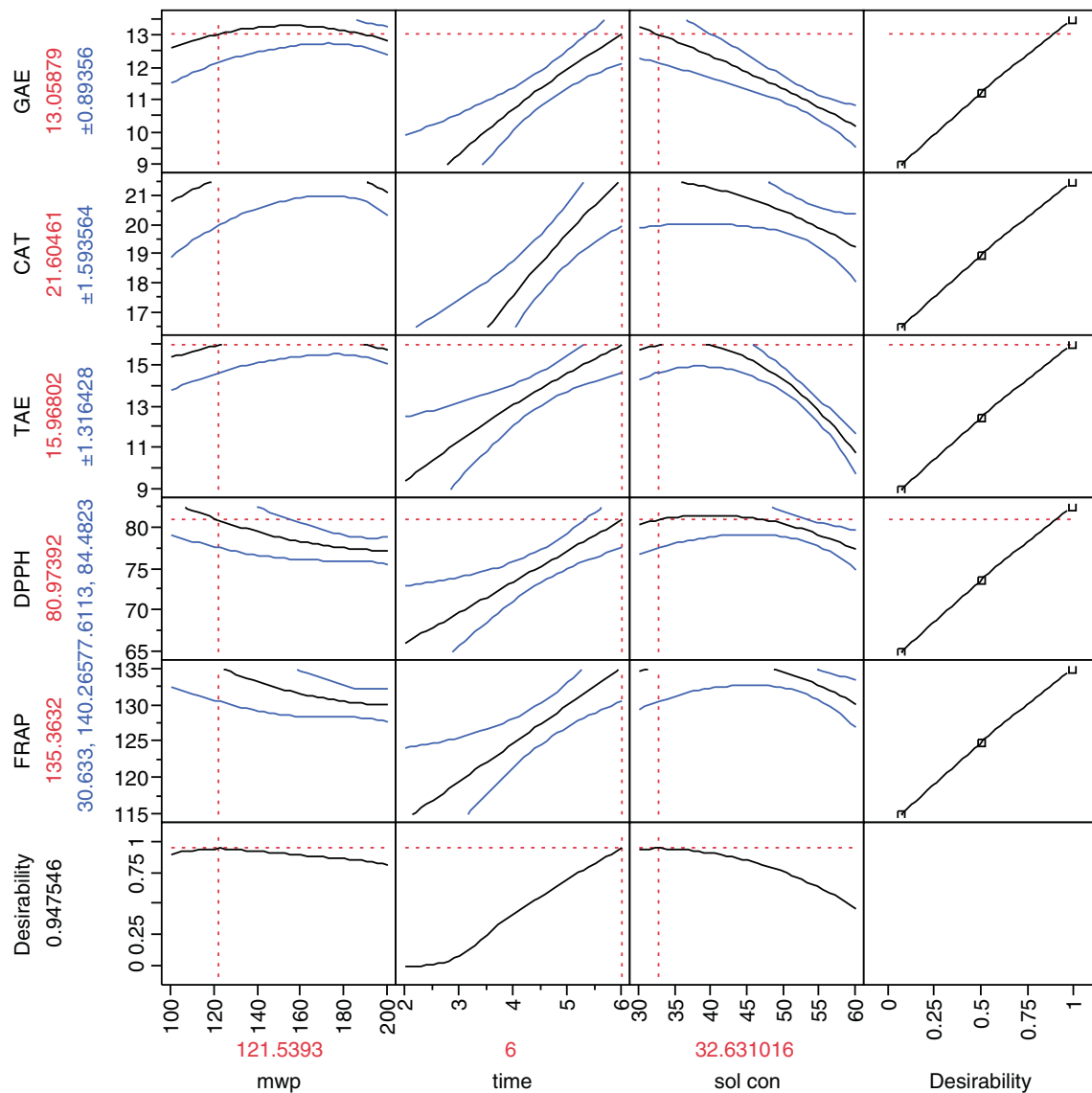


Fig. 7 Overall optimum conditions of microwave-assisted extraction of phenolic antioxidants from grape seed

cyanidin-based anthocyanin pigments; it was found that methoxylation and hydroxylation of the B ring had predictable effect on color. Katalinic et al. (2004) compared the reducing power (FRAP) and antioxidant effectiveness of red wines with white wines. It was found that average FRAP of red wines was almost tenfold higher than the average FRAP of white wines, which was mainly due to the presence of anthocyanin pigments and catechins that could significantly participate in the antioxidant power of red wine. This indicates that grape seed antioxidants make significant contribution to the color of the extract. Thus chromatic parameters can be a helpful means for comparison and quick assessment of phenolic content in grape seed extract. Further studies are required in this area to understand the concepts of coloring compounds and antioxidant activity.

#### Optimum Conditions for Maximum Extraction from Grape Seeds

It is relatively simple to determine the exact optimum conditions for single response using RSM (Bezerra et al. 2008, Lundstedt et al. 1998). The predicted optimal conditions for obtaining maximum total phenols were: 170 W, 6 min, and 30% ethanol concentration for  $13.5 \pm 0.48$  mg GAE/g dw GS with a desirability of 0.982; 168 W, 6 min, and 30% EtOH for  $22 \pm 0.86$  mg CAT/g dw GS with a desirability of 0.997; and 160 W, 6 min, and 33.7% ethanol for  $16.1 \pm 0.56$  mg TAE/g dw GS with a desirability of 0.992. The maximum antioxidant activity was found at 100 W, 6 min, and 39.4% ethanol concentration for DPPH providing 86.5% DPPH inhibition with desirability of 0.997, and 100 W, 6 min, and 39.2% ethanol concentration for FRAP giving 143 ( $\mu\text{M}$  of ascorbic acid eq)/g dw GS with a desirability of 0.998.

When various responses have to be considered at the same time, it is necessary to find optimal compromises between the total number of response taken into account. Thus Derringer function or desirability function ( $d$ ) is the most important and most currently used multicriteria methodology in optimization procedures (Bezerra et al. 2008). In order to maximize the response of the dependent variables, a numerical optimization was carried by statistical means which provided the overall optimum conditions of extraction of antioxidants from grape seed at 121 W, 6 min, and 32.6% ethanol concentration with desirability of 0.947 (Fig. 7). The numerical optimization provided the maximum predicted total phenols in terms of various equivalents as  $13 \pm 0.89$  mg GAE/g dw grape seed (GS),  $21.6 \pm 1.59$  mg CAT/g dw GS, and  $15.9 \pm 1.32$  mg TAE/g dw GS, and the values for antioxidant activity were 80.9% DPPH inhibition and  $135 \mu\text{M}$  of ascorbic acid equivalents/g dw GS for FRAP.

The application of desirability function in optimization of multiple responses in extraction procedures makes it more

efficient and economical. Thus the maximum desirability in terms of total phenols was obtained for catechin ( $d=0.997$ ) which was the basic monomeric unit present in grape seed phenols. In terms of antioxidant activity, the desirability function for both DPPH and FRAP was almost same. Katalinic et al. (2004) studied the antioxidant effectiveness of grape wines in comparison with (+)-catechin and demonstrated significant antioxidant capacity with FRAP assay. FRAP is a very rapid and simple method and is very convenient for screening large number of samples (Katalinic et al. 2004). Thus a multiple response optimization was done for total phenols in terms of catechin, and FRAP resulted  $21.6 \pm 1.51$  mg CAT/g dw GS and  $135 \mu\text{M}$  of ascorbic acid equivalents/g dw GS with a maximum desirability of 0.991 at 122 W, 6 min, and 33.7% ethanol concentration.

#### Conclusion

Thus the principles of multiple response optimization using desirability function can be applied to the development of extraction procedures that demand a search for optimal conditions for a set of response simultaneously. In our study, we were able to extract maximum phenolic antioxidant with lower solvent consumption and extraction time by microwave-assisted extraction when compared with results reported in the literature. With the increasing consumer interest in functional foods, food companies are looking to incorporate new bioactive ingredients to improve health. Thus grape seed which is considered as a winery waste can be used as a high value functional food ingredient as substituted or alternative antioxidant to the synthetic antioxidants. This will in turn increase the economic value of the grape seed obtained from grape pomace.

**Acknowledgments** We thank the Dept. of Foreign Affairs and International Trade, Canada for providing the Canadian Commonwealth Scholarship, 2010. Thanks to Simona Nemes and Ashutosh Singh, Dept. of Bioresource Engineering, McGill University for their technical expertise and support.

#### References

- Abdeshahian, P., Samat, N., & Yusoff, W. M. W. (2010). Utilization of palm kernel cake for production of beta-glucosidase by *Aspergillus niger* FTCC 5003 in solid substrate fermentation using an aerated column bioreactor. *Biotechnology*, 9(1), 17–24.
- Al-Awwadi, N. A., Araiz, C., Bornet, A., Delbosc, S., Cristol, J. P., Linck, N., et al. (2005). Extracts enriched in different polyphenolic families normalize increased cardiac NADPH oxidase expression while having differential effects on insulin resistance, hypertension, and cardiac hypertrophy in high-fructose-fed rats. *Journal of Agricultural and Food Chemistry*, 53(1), 151–157.
- Bail, S., Stuebiger, G., Krist, S., Unterweger, H., & Buchbauer, G. (2008). Characterisation of various grape seed oils by volatile



- compounds, triacylglycerol composition, total phenols and antioxidant capacity. *Food Chemistry*, 108(3), 1122–1132.
- Ballard, T. S., Mallikarjunan, P., Zhou, K., & O'Keefe, S. F. (2009). Optimizing the extraction of phenolic antioxidants from peanut skins using response surface methodology. *Journal of Agricultural and Food Chemistry*, 57(8), 3064–3072.
- Baoshan Sun, M., & Spranger, I. (2005). Review: quantitative extraction and analysis of grape and wine proanthocyanidins and stilbenes. *Ciência Tec Vitiv*, 20(2), 59–89.
- Benzie, I. F. F., & Strain, J. (1996). The ferric reducing ability of plasma (FRAP) as a measure of "antioxidant power": the FRAP assay. *Analytical Biochemistry*, 239(1), 70–76.
- Bezerra, M. A., Santelli, R. E., Oliveira, E. P., Villar, L. S., & Escaleira, L. A. (2008). Response surface methodology (RSM) as a tool for optimization in analytical chemistry. *Talanta*, 76(5), 965–977.
- Buci Koji, A., Planini, M., Tomas, S., Jakobek, L., & Šeruga, M. (2009). Influence of solvent and temperature on extraction of phenolic compounds from grape seed, antioxidant activity and colour of extract. *International Journal of Food Science and Technology*, 44(12), 2394–2401.
- Cacace, J., & Mazza, G. (2003). Mass transfer process during extraction of phenolic compounds from milled berries. *Journal of Food Engineering*, 59(4), 379–389.
- Canals, R., Llaudy, M., Valls, J., Canals, J., & Zamora, F. (2005). Influence of ethanol concentration on the extraction of color and phenolic compounds from the skin and seeds of Tempranillo grapes at different stages of ripening. *Journal of Agricultural and Food Chemistry*, 53(10), 4019–4025.
- Castillo, J., Benavente-Garcia, O., Lorente, J., Alcaraz, M., Redondo, A., Ortuno, A., et al. (2000). Antioxidant activity and radioprotective effects against chromosomal damage induced in vivo by X-rays of flavan-3-ols (Procyanidins) from grape seeds (*Vitis vinifera*): Comparative study versus other phenolic and organic compounds. *Journal of Agricultural and Food Chemistry*, 48(5), 1738–1745.
- Chen, S., & Spiro, M. (1994). Study of microwave extraction of essential oil constituents from plant materials. *The Journal of Microwave Power and Electromagnetic Energy*, 29(4), 231–241.
- Cheyrier, V., & Rigaud, J. (1986). HPLC separation and characterization of flavonols in the skins of *Vitis vinifera* var. Cinsault. *American Journal of Enology and Viticulture*, 37(4), 248.
- Craveiro, A., Matos, F., Alencar, J., & Plumel, M. (1989). Microwave oven extraction of an essential oil. *Flavour and Fragrance Journal*, 4(1), 43–44.
- Dandekar, D. V., & Gaikar, V. (2002). Microwave assisted extraction of curcuminoids from *Curcuma longa*. *Separation Science and Technology*, 37(11), 2669–2690.
- Deighton, N., Brennan, R., Finn, C., & Davies, H. V. (2000). Antioxidant properties of domesticated and wild *Rubus* species. *Journal of the Science of Food and Agriculture*, 80(9), 1307–1313.
- Del Bas, J. M., Fernández-Larrea, J., Blay, M., Ardévol, A., Salvadó, M. J., Arola, L., et al. (2005). Grape seed procyanidins improve atherosclerotic risk index and induce liver CYP7A1 and SHP expression in healthy rats. *The FASEB Journal*, 19(3), 479.
- FAOSTAT. (2011). World production quantity of grapes: 2009. Food and Agricultural Organization of the United Nations. FAO Statistics Division 2011. Available at: <http://faostat.fao.org/>. Accessed 04 October 2011.
- Freitas, V. A. P., Glories, Y., Bourgeois, G., & Vitry, C. (1998). Characterisation of oligomeric and polymeric procyanidins from grape seeds by liquid secondary ion mass spectrometry. *Phytochemistry*, 49(5), 1435–1441.
- Ganzler, K., Szinai, I., & Salgo, A. (1990). Effective sample preparation method for extracting biologically active compounds from different matrices by a microwave technique. *Journal of Chromatography. A*, 520, 257–262.
- Garcia-Marino, M., Rivas-Gonzalo, J. C., Ibanez, E., & Garcia-Moreno, C. (2006). Recovery of catechins and proanthocyanidins from winery by-products using subcritical water extraction. *Analytica Chimica Acta*, 563(1–2), 44–50.
- Ghafoor, K., Choi, Y. H., Jeon, J. Y., & Jo, I. H. (2009). Optimization of ultrasound-assisted extraction of phenolic compounds, antioxidants, and anthocyanins from grape (*Vitis vinifera*) seeds. *Journal of Agricultural and Food Chemistry*, 57(11), 4988–4994.
- Gokturk Baydar, N., Ozkan, G., & Yasar, S. (2007). Evaluation of the antiradical and antioxidant potential of grape extracts. *Food Control*, 18(9), 1131–1136.
- Gómez-Alonso, S., García-Romero, E., & Hermosín-Gutiérrez, I. (2007). HPLC analysis of diverse grape and wine phenolics using direct injection and multidetection by DAD and fluorescence. *Journal of Food Composition and Analysis*, 20(7), 618–626.
- Guo, L., Wang, L. H., Sun, B., Yang, J. Y., Zhao, Y. Q., Dong, Y. X., et al. (2007). Direct in vivo evidence of protective effects of grape seed procyanidin fractions and other antioxidants against ethanol-induced oxidative DNA damage in mouse brain cells. *Journal of Agricultural and Food Chemistry*, 55(14), 5881–5891.
- Haaland, P. D. (1989). *Experimental design in biotechnology*. Boca Raton: CRC.
- Harman, D. (1995). Role of antioxidant nutrients in aging: overview. *Age*, 18(2), 51–62.
- Hong, N., Yaylayan, V. A., Raghavan, G. S., Paré, J. R., & Bélanger, J. M. (2001). Microwave-assisted extraction of phenolic compounds from grape seed. *Natural Product Letters*, 15(3), 197.
- Jayaprakasha, G. K., & Jaganmohan Rao, L. (2000). Phenolic constituents from lichen *Parmotrema stippeum* (Nyl.) Hale and their antioxidant activity. *Zeitschrift für Naturforschung*, 55, 1018–1022.
- Jayaprakasha, G., Singh, R., & Sakariah, K. (2001). Antioxidant activity of grape seed (*Vitis vinifera*) extracts on peroxidation models in vitro. *Food Chemistry*, 73(3), 285–290.
- Jayaprakasha, G., Selvi, T., & Sakariah, K. (2003). Antibacterial and antioxidant activities of grape (*Vitis vinifera*) seed extracts. *Food Research International*, 36(2), 117–122.
- Katalinic, V., Milos, M., Modun, D., Musi, I., & Boban, M. (2004). Antioxidant effectiveness of selected wines in comparison with (+)-catechin. *Food Chemistry*, 86(4), 593–600.
- Kim, S. Y., Jeong, S. M., Park, W. P., Nam, K., Ahn, D., & Lee, S. C. (2006). Effect of heating conditions of grape seeds on the antioxidant activity of grape seed extracts. *Food Chemistry*, 97(3), 472–479.
- Koundouras, S., Marinos, V., Gkoulioti, A., Kotseridis, Y., & van Leeuwen, C. (2006). Influence of vineyard location and vine water status on fruit maturation of nonirrigated cv. Agiorgitiko (*Vitis vinifera* L.). Effects on wine phenolic and aroma components. *Journal of Agricultural and Food Chemistry*, 54(14), 5077–5086.
- Lee, J., Koo, N., & Min, D. (2004). Reactive oxygen species, aging, and antioxidative nutraceuticals. *Comprehensive Reviews in Food Science and Food Safety*, 3(1), 21–33.
- Li, H., Wang, X., Li, P., Li, Y., & Wang, H. (2008). Comparative study of antioxidant activity of grape (*Vitis vinifera*) seed powder assessed by different methods. *Journal of Food and Drug Analysis*, 16, 67–73.
- Lundstedt, T., Seifert, E., Abramo, L., Thelin, B., Nystrom, A., Pettersen, J., et al. (1998). Experimental design and optimization. *Chemometrics and Intelligent Laboratory Systems*, 42(1–2), 3–40.
- Maier, T., Schieber, A., Kammerer, D. R., & Carle, R. (2009). Residues of grape (*Vitis vinifera* L.) seed oil production as a valuable source of phenolic antioxidants. *Food Chemistry*, 112(3), 551–559.
- Mandic, A. I., Ilas, S. M., Etkovi, G. S., Anadanovi-Brunet, J. M., & Tumbas, V. T. (2008). Polyphenolic composition and antioxidant activities of grape seed extract. *International Journal of Food Properties*, 11(4), 713–726.

- Mayer, R., Stecher, G., Wuerzner, R., Silva, R. C., Sultana, T., Trojer, L., et al. (2008). Proanthocyanidins: target compounds as antibacterial agents. *Journal of Agricultural and Food Chemistry*, 56(16), 6959–6966.
- Mazza, G., & Francis, F. (1995). Anthocyanins in grapes and grape products. *Critical Reviews in Food Science and Nutrition*, 35(4), 341–371.
- Meyer, A. S., Yi, O. S., Pearson, D. A., Waterhouse, A. L., & Frankel, E. N. (1997). Inhibition of human low-density lipoprotein oxidation in relation to composition of phenolic antioxidants in grapes (*Vitis vinifera*). *Journal of Agricultural and Food Chemistry*, 45(5), 1638–1643.
- Nair, V. D. P., Dairam, A., Agbonon, A., Arnason, J., Foster, B., & Kanfer, I. (2007). Investigation of the antioxidant activity of African potato (*Hypoxis hemerocallidea*). *Journal of Agricultural and Food Chemistry*, 55(5), 1707–1711.
- Nemes, S. M., & Orsat, V. (2009). Microwave-assisted extraction of secoisolariciresinol diglucoside—Method development. *Food and Bioprocess Technology* 1-9.
- Nemes, S. M., & Orsat, V. (2010). Screening the experimental domain for the microwave-assisted extraction of secoisolariciresinol diglucoside from flaxseed prior to optimization procedures. *Food and Bioprocess Technology*, 3(2), 300–307.
- Pallaroni, L., Von Holst, C., Eskilsson, C., & Björklund, E. (2002). Microwave-assisted extraction of zearalenone from wheat and corn. *Analytical and Bioanalytical Chemistry*, 374(1), 161–166.
- Pan, X., Liu, H., Jia, G., & Shu, Y. Y. (2000). Microwave-assisted extraction of glycyrrhizic acid from licorice root. *Biochemical Engineering Journal*, 5(3), 173–177.
- Pan, X., Niu, G., & Liu, H. (2003). Microwave-assisted extraction of tea polyphenols and tea caffeine from green tea leaves. *Chemical Engineering and Processing*, 42(2), 129–133.
- Pan, Y., Wang, K., Huang, S., Wang, H., Mu, X., He, C., et al. (2008). Antioxidant activity of microwave-assisted extract of longan (*Dimocarpus Longan* Lour.) peel. *Food Chemistry*, 106(3), 1264–1270.
- Pinent, M., Blay, M., Blade, M., Salvado, M., Arola, L., & Ardevol, A. (2004). Grape seed-derived procyanidins have an antihyperglycemic effect in streptozotocin-induced diabetic rats and insulinomimetic activity in insulin-sensitive cell lines. *Endocrinology*, 145(11), 4985.
- Pokorn, J. (1991). Natural antioxidants for food use. *Trends in Food Science & Technology*, 2, 223–227.
- Puiggròs, F., Llopiz, N., Ardévol, A., Bladé, C., Arola, L., & Salvadó, M. J. (2005). Grape seed procyanidins prevent oxidative injury by modulating the expression of antioxidant enzyme systems. *Journal of Agricultural and Food Chemistry*, 53(15), 6080–6086.
- Ramchandani, A. G., Chettiyar, R. S., & Pakhale, S. S. (2010). Evaluation of antioxidant and anti-initiating activities of crude polyphenolic extracts from seedless and seeded Indian grapes. *Food Chemistry*, 119(1), 298–305.
- Singh, A., Sabally, K., Kubow, S., Donnelly, D. J., Garipey, Y., Orsat, V., et al. (2011). Microwave-assisted extraction of phenolic antioxidants from potato peels. *Molecules*, 16(3), 2218–2232.
- Spranger, I., Sun, B., Mateus, A. M., Freitas, V., & Ricardo-da-Silva, J. M. (2008). Chemical characterization and antioxidant activities of oligomeric and polymeric procyanidin fractions from grape seeds. *Food Chemistry*, 108(2), 519–532.
- Stintzing, F. C., Stintzing, A. S., Carle, R., Frei, B., & Wrolstad, R. E. (2002). Color and antioxidant properties of cyanidin-based anthocyanin pigments. *Journal of Agricultural and Food Chemistry*, 50(21), 6172–6181.
- Sun, Y., Liao, X., Wang, Z., Hu, X., & Chen, F. (2007). Optimization of microwave-assisted extraction of anthocyanins in red raspberries and identification of anthocyanin of extracts using high-performance liquid chromatography–mass spectrometry. *European Food Research and Technology*, 225(3), 511–523.
- Terra, X., Valls, J., Vitrac, X., Mérrillon, J. M., Arola, L., Ardévol, A., et al. (2007). Grape-seed procyanidins act as anti-inflammatory agents in endotoxin-stimulated RAW 264.7 macrophages by inhibiting NFκB signaling pathway. *Journal of Agricultural and Food Chemistry*, 55(11), 4357–4365.
- Thostenson, E., & Chou, T. W. (1999). Microwave processing: fundamentals and applications. *Composites Part A: Applied Science and Manufacturing*, 30(9), 1055–1071.
- Wang, S., Chen, F., Wu, J., Wang, Z., Liao, X., & Hu, X. (2007). Optimization of pectin extraction assisted by microwave from apple pomace using response surface methodology. *Journal of Food Engineering*, 78(2), 693–700.
- Wang, Y., Xi, G. S., Zheng, Y. C., & Miao, F. S. (2010). Microwave-assisted extraction of flavonoids from Chinese herb *Radix puerariae* (Ge Gen). *Journal of Medicinal Plant Research*, 4(4), 304–308.
- Wrolstad, R. E., Durst, R. W., & Lee, J. (2005). Tracking color and pigment changes in anthocyanin products. *Trends in Food Science & Technology*, 16(9), 423–428.
- Yamakoshi, J., Saito, M., Kataoka, S., & Tokutake, S. (2002). Procyanidin-rich extract from grape seeds prevents cataract formation in hereditary cataractous (ICR/f) rats. *Journal of Agricultural and Food Chemistry*, 50(17), 4983–4988.
- Zumbo, B., & Harwell, M. (1999). The methodology of methodological research: Analyzing the results of simulation experiments (Paper No. ESQBS-99-2). Prince George, BC: University of Northern British Columbia Edgeworth Laboratory for Quantitative Behavioral Science.



This article was downloaded by: [McGill University Library]

On: 29 July 2013, At: 10:23

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Physics and Chemistry of Liquids: An International Journal

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/gpch20>

### Variation of dielectric properties of aqueous solutions of ethanol and acids at various temperatures with low acid concentration levels

Winy Routray<sup>a</sup> & Valerie Orsat<sup>a</sup>

<sup>a</sup> Bioresource Engineering Department, Macdonald Campus , McGill University , Ste-Anne-de-Bellevue , QC , Canada , H9X 3V9  
Published online: 26 Jul 2013.

To cite this article: Physics and Chemistry of Liquids (2013): Variation of dielectric properties of aqueous solutions of ethanol and acids at various temperatures with low acid concentration levels, Physics and Chemistry of Liquids: An International Journal, DOI: 10.1080/00319104.2013.812022

To link to this article: <http://dx.doi.org/10.1080/00319104.2013.812022>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms &

Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

## Variation of dielectric properties of aqueous solutions of ethanol and acids at various temperatures with low acid concentration levels

Winy Routray\* and Valerie Orsat

*Bioresource Engineering Department, Macdonald Campus, McGill University,  
Ste-Anne-de-Bellevue, QC, Canada H9X 3V9*

*(Received 28 April 2013; final version received 19 May 2013)*

Microwave-assisted extraction is gaining in popularity for the production of plant extracts and the dielectric properties determine the behaviour of different materials in a microwave environment; hence the measurement of the dielectric properties of different solvents used for analytical purposes and the interpretation of their heating rate can help in minimising the time and solvent consumption. In this study, the dielectric constant and dielectric loss of different combinations (generated using central composite design (CCD)) of ethanol concentrations and acid (HCl and citric acid) at different temperature levels were measured at frequencies varying from 0.5 GHz to 6 GHz, and the dissipation factor and depth of penetration were calculated for the respective cases, which were analysed for the interpretation of heating rates of the different solvent combinations. Overall, for all the responses (observed and calculated), ethanol concentration levels and temperature levels were found to be the most significant factors; it was interpreted and observed that the heating rate increases with an increase in ethanol concentration and decreases with increasing temperature. In general, the dissipation factor was interpreted to increase with increase in acid concentration (observed from screening analysis); however, citric acid concentration was found to have an insignificant effect on the dissipation factor (as interpreted from CCD with response surface methodology (CCD-RSM)), and hence was found to have insignificant contribution to the heating rate at lower concentrations. In case of CCD-RSM analysis of the values of dissipation factor for the combinations of ethanol with HCl, the effect of acid concentration on the dissipation factor was found to be significant; however, it was still found to be negligible when the solvent combinations were heated in the microwave reactor.

**Keywords:** dielectric constant; dielectric loss; hydrochloric acid; citric acid

### 1. Introduction

Microwave frequencies 915, 2450 and 5800 MHz have been permitted by Federal Communications Commission (FCC) for different industrial applications, and are widely used in post-harvest, food processing operations and analytical studies [1]. However, most of the commercial microwave processing equipments are designed for operation at 2450 MHz [1]. Microwave-assisted extraction is another application of microwaves which has been increasingly applied for the extraction of nutraceutical compounds [2–5]. The major advantages of microwave-assisted extraction are higher rate of heating, which leads to a faster extraction rate, and lower solvent consumption. Ethanol is one of the common solvents used for the extraction of nutraceuticals, and the combination of ethanol with acids is a familiar solvent combination used in the extraction of several organic

---

\*Corresponding author. Email: routrayw@yahoo.com

compounds [6–8], where acids are added to increase the rate of deterioration of cellular structure leading to increased diffusion of target compounds into the surrounding solvent.

The dissemination of microwaves in materials is governed by Maxwell's equations, which relate the electric field and magnetic field vectors to other property parameters such as permittivity and permeability. Permittivity, which characterises the interaction between microwave electric field and material, is generally expressed as the relative complex permittivity (permittivity of material with respect to permittivity of free space). Relative complex permittivity can be expressed in terms of dielectric properties, which determine the response of different solvents and materials exposed to microwave. The dielectric properties of materials considered are the dielectric constant and dielectric loss, which are related to the relative complex permittivity through the expression:

$$\epsilon_r = \epsilon' - j\epsilon'' \quad (1)$$

where  $\epsilon_r$  is the relative complex permittivity,  $\epsilon'$  is the dielectric constant,  $\epsilon''$  is the dielectric loss, and  $j = \sqrt{-1}$ . The dielectric constant is associated with the ability of the material to store electrical energy in an electromagnetic field and the dielectric loss factor reflects the capacity of conversion of the electromagnetic energy into thermal energy [1]. Other parameters such as relaxation time (time taken to achieve the randomised orientation state, when the electric field applied to a polar solution is switched off) and loss tangent or dissipation factor ( $\tan \delta$ ) are also used to explain the behaviour of materials and solvents in the microwave environment. For polar liquids with single relaxation time, complex permittivity can be expressed by the Debye equation which relates it to relaxation time. The dielectric constant and dielectric loss can also be expressed in terms of relaxation time using the Debye equation [9]. In the last few decades, there have been many studies on the dielectric behaviour of solutions with multiple chemical components [10,11], which also include binary solutions, such as methanol and ethanol [12,13] and their individual combinations with other compounds [14]. During these studies, the binary systems were found to obey the Debye model, which could be used to predict the dielectric behaviour of the solvent combination. Many of these studies concentrated on the dielectric behaviour of solvent mixtures, and investigated parameters of solvent mixtures including static dielectric constant, dielectric relaxation time and high-frequency dielectric constant, that are related to the corresponding properties of the pure solvents through appropriate dielectric mixing rules, and were used to calculate the relative complex permittivity of the solvent mixture [12]. However, during a recent study by Yang et al. [12], a modification of the Debye equation for the determination and explanation of dielectric behaviour of mixture systems has been emphasised. Different direct measurement methods for determination of dielectric properties, and subsequent explanation and prediction of solvent or material behaviour in case of microwave application are recently attracting more attention [12,15]. Various methods for the measurement of dielectric constant and dielectric loss have been reported by Datta et al. [16].

The study of the dielectric properties of solvents is helpful in predicting the power dissipation factor, depth of penetration and heating rate [9]. For materials with comparable dielectric constant, the dissipation factor is a useful parameter for comparing the efficiency of conversion of microwave energy into thermal energy, and for compounds with comparable chemical and physical characteristics, it is also useful in comparing the consequent heating rates. As microwaves travel in a material, they get attenuated.

Penetration depth is the distance over which 63% of the power is dissipated and it depends on the properties of the material; hence information regarding penetration depth helps in assessing 'relative microwave absorbing characteristics of the materials' [17,18]. Frequency of the applied microwave, temperature of the solution, molecular mass of the solvent compounds, compound volumes and nature of bonds within the molecules and between the molecules are some of the important properties affecting the dielectric properties of a final solution mixture. The microwave frequencies of our interest are the three FCC-permitted frequencies and especially 2.45 GHz, used for extraction purposes. The extent to which the substances are exposed to microwave and are able to couple with the microwave radiation is controlled by the magnitudes of dielectric constant and dielectric loss, and their property of frequency dependence, because of which they are the essential parameters for interpretation of the dielectric heating phenomenon [9]. Alcohols are capable of forming hydrogen bonds similar to water; however, to a lesser extent. Ethanol is one of the alcohols which have been reported to have proper relaxation properties, which enable it to couple effectively with microwave at 2.45 GHz and, hence, has been reported as an effective solvent for dielectric heating [9]. According to the information collected from different reports, there are no studies on the dielectric properties of alcohol–acid combinations, which are quite common solvent combinations in the extraction of nutraceuticals, especially flavonoid groups. Citric acid and hydrochloric acid are some of the common acids used for the purpose, where hydrochloric acid is a strong acid and citric acid is a weaker acid (the pKa value for HCl is much lower than that of citric acid, which implies the comparative acidity level), which were selected for this particular study. These acids have been found to be polar compounds and for polar molecules with low molecular weight; the relaxation process occurs in the microwave frequency range [9], which implies that the above-mentioned acids can be expected to be potential solvents for dielectric extraction applications.

This study focussed on the determination of the effect of combinations of different acid and alcohol concentrations at different temperatures on dielectric constant, dielectric loss, loss tangent (dissipation factor) and depth of penetration of acid and alcohol mixtures. Variation of dielectric constant, dielectric loss and dissipation factor with respect to different frequencies was studied. The effect of individual factors (different temperature levels, acid and alcohol concentrations) on dielectric constant, dielectric loss, dissipation factor and depth of penetration at 2450 MHz, was analysed using screening, central composite design (CCD) and response surface methodology (RSM). To study the effect of lower molarity, HCl and citric acid were used for the experiments. Finally, the variation of temperature (basically rate of increase of temperature) with the time of application of microwave in a microwave reactor was recorded for some of the combinations of alcohol and acid concentration levels and analysed.

## 2. Experimental methods

### 2.1. Materials

Anhydrous ethyl alcohol (Commercial alcohols, Brampton, Ontario, Canada; 100% v/v), HCl (Fisher Scientific, Fair Lawn, New Jersey, USA; 12.1 N, specific gravity 1.19) and citric acid (Sigma Aldrich, St. Louis, MO, USA; 99.6%) were used to prepare the respective solutions with water (HPLC grade water prepared using Simplicity™ Water Purification System (Millipore, Billerica, MA, USA)) at room temperature (293.15–298.15 K) with different concentrations as per the experimental design.



## 2.2. Statistical design and analysis

CCD with uniform precision was used as the experimental design for this study. Three factors, namely temperature (X1), acid concentration (X2) and ethanol concentration (X3), each at five levels (-1.682, -1, 0, 1, 1.682, Table 1) were used for preparing the experimental design with a total of 20 treatment combinations, consisting of 8 factorial, 6 axial and 6 central (for increase of uniform precision) point combinations (Table 2). JMP software version 8 (SAS Institute Inc., Cary, NC, USA) was used for the experimental design and analysis. Separate CCDs were planned for each acid, combined with different alcohol concentrations (9.73–70.3% v/v) and temperature levels (24.8–65.2°C or 297.95–338.35 K). As mentioned before, for identical molarity range (0.496–1.505 mol/dm<sup>3</sup>) of citric and hydrochloric acids, separate CCDs were planned for each acid (Table 1). The statistical design was prepared taking temperature in °C and acid concentration in molarity (M, mol/dm<sup>3</sup>) rather than in Kelvin and in molality, respectively. The corresponding experimental design, the observations, regression equations and response surface plots will be different if the temperature is presented in Kelvin and concentration in molality, respectively, because of the relevant conversions. However, the corresponding temperatures in Kelvin and concentration in molality have been presented in Table 2 and this particular fact has been mentioned several times in the manuscript.

Table 2 shows the different treatment combinations (where one is the centre point which has 6 replicates, for a total of 20 treatment combinations). Graphs were plotted for different parameters, such as dielectric constant, dielectric loss and dissipation factor with respect to microwave frequencies in the range of 0.5–6 GHz, to observe and analyse the patterns of variation of these factors with the changing frequency.

The factorial parts of the experimental design for both the acids were used to generate a screening design for analysing and choosing the more influential acid and to observe the individual and interactive effects of the different tested factors on dielectric constant, dielectric loss and dissipation factor at 2450 MHz microwave frequency. There were four factors considered for this analysis, one qualitative factor, i.e. types of acids (HCl and citric acid), three quantitative variables, which were ethanol concentrations (22% and 58% v/v), temperature levels (33°C and 57°C or 306.15–330.15 K) and acid concentrations (0.7 and 1.3 mol/dm<sup>3</sup>); all with two levels. There were 16 treatment combinations and three replicates of each were considered for these analyses, so in total there were 48 treatment combinations.

In the later part of the analysis of this study, the data collected for all the 20 treatment combinations of CCD were analysed, and individual (linear), quadratic and combined (bilinear) effects of various factors on the dissipation factor were interpreted and

Table 1. Independent variables and their coded and actual values used for the central composite design.

Independent variables	Unit	Coded levels				
		-1.682	-1	0	1	1.682
Temperature (Temp)	°C	24.8	33	45	57	65.2
Ethanol concentration (EtOH)	% v/v	9.72	22	40	58	70.3
Acid concentration (M) (citric acid/hydrochloric acid) (CA/HCl)	mol/dm <sup>3</sup>	0.496	0.7	1	1.3	1.505

Note: The design was prepared considering temperature in °C and acid concentration in molarity; however, design points will be different if it is created in Kelvin and molality, respectively.

Table 2. Central composite showing the different combination of the temperature levels, different alcohol and different types of acids with different concentrations.

Experimental combination no.	Temperature		Ethanol concentration (% v/v)	Molarity (M, mol/dm <sup>3</sup> )	Acid concentration		CA (mol/Kg)	Factorial points	
	°C	K			HCl (mol/Kg)				
1	33	306.15	22	0.7	0.687	0.664	0.664	Factorial points	
2	33	306.15	22	1.3	1.274	1.182	1.182		
3	33	306.15	58	0.7	0.687	0.664	0.664		
4	33	306.15	58	1.3	1.274	1.182	1.182		
5	57	330.15	22	0.7	0.687	0.664	0.664		
6	57	330.15	22	1.3	1.274	1.182	1.182		
7	57	330.15	58	0.7	0.687	0.664	0.664		
8	57	330.15	58	1.3	1.274	1.182	1.182		
9	24.8	297.95	40	1	0.985	0.929	0.929		Axial points
10	65.2	338.35	40	1	0.985	0.929	0.929		
11	45	318.15	9.73	1	0.985	0.929	0.929		
12	45	318.15	70.3	1	0.985	0.929	0.929		
13	45	318.15	40	0.496	0.491	0.477	0.477		
14	45	318.15	40	1.505	1.47	1.349	1.349		
15-20	45	318.15	40	1	0.985	0.929	0.929		Centre points

Note: Experiment was designed considering Temperature in °C and acid concentration in molarity and not molality.

evaluated, and regression models for the dissipation factor were generated for different acid and ethanol combinations separately, at different temperature levels, whose validity was analysed using ANOVA. The response function dissipation factor ( $Y$ ) was partitioned into linear, quadratic and interactive components:

$$Y = \beta_0 + \sum_{i=1}^k B_i X_i + \sum_{i=1}^k B_{ii} X_i^2 + \sum_{i>j}^k B_{ij} X_i X_j \quad (2)$$

where  $\beta_0$  denotes the model intercept;  $B_i$ ,  $B_{ii}$ ,  $B_{ij}$  represent the regression coefficients of the linear, quadratic and interactive (bilinear) effects, respectively;  $X_i$  and  $X_j$  are the coded independent variables, and  $k$  is the number of independent variables or factors, which is 3 in this case. Tables of the analysis of variance were generated, the regression coefficients were determined, and the significances of all terms in the polynomial were evaluated statistically based on the  $F$ -value at the probability ( $p$ ) of 0.001, 0.01 or 0.05. Three-dimensional figures (response surfaces) were also plotted with 2 factors (in each case 2 factors out of the 3 tested factors which included ethanol concentration, temperature level and acid molarity) on  $X$  and  $Y$  axes and response on  $Z$  axis to demonstrate the perceptible individual and combined effects of 2 factors, on the response, dissipation factor.

### 2.3. Dielectric measurement

The mixtures of ethanol with different concentrations and acids with different molarity levels were prepared from the standard solutions (corresponding compounds mixed or diluted in water) and stored, and just before the dielectric measurements, mixed solutions of acid with ethanol were prepared. The proportion of alcohol solution to acid solution in terms of volume was fixed at 99:1. The total volume of acid and alcohol solution was 20 ml (19.8 ml alcohol solution and 0.2 ml acid solution); graduated glass tubes were used for preparing the solution and for subsequent measurement. A low acid concentration range was chosen in this study. Dielectric measurements were made by a dielectric measurement system (Agilent S Parameter Network Analyser, 8722ES, Agilent Technologies, Santa Clara, CA, USA). Open-ended high-temperature (flat) probe (85070D, Agilent Technologies, USA) was used along with the ECal module (N4693A, Agilent technologies, USA) for the measurement of the dielectric constant ( $\epsilon'$ ) and dielectric loss ( $\epsilon''$ ), from which the dissipation factor was calculated. This method is more advantageous than many of the transmission line measurement techniques. This method calculates the dielectric properties from the 'phase and amplitude of the reflected signal at the end of an open-ended coaxial line.' [19] The sample solutions were heated to the temperature levels as planned by the CCD using an Isotemp heater (2001FS, Thermofisher Scientific, Waltham, MA, USA) and the probe was inserted into the solutions. The manipulations for the measurement were made very fast to ensure that the temperature was only altered in the range of  $\pm 0.05^\circ\text{C}$  or K which was not found to affect the dielectric properties. The range of microwave frequency used for the study was 0.5–6 GHz, which included all the frequencies used for commercial applications of microwaves. Measurements were obtained for three replicates of the treatment combinations of CCD with citric acid and hydrochloric acid for the range of molarity of  $0.496 \text{ mol/dm}^3$  to  $1.505 \text{ mol/dm}^3$ . High repeatability was observed during this experimental study.

Dissipation factor was calculated using the equation

$$\tan \delta = \frac{\epsilon''}{\epsilon'} \quad (3)$$

and depth of penetration was calculated using the following equation [16]

$$\delta_p = \frac{\lambda_0}{2\pi\sqrt{2\epsilon'}} \left( \sqrt{1 + (\epsilon''/\epsilon')^2} - 1 \right)^{-1/2} \quad (4)$$

where,  $\delta_p$  is the penetration depth (m) and  $\lambda_0$  is the wavelength of microwaves in free space (0.122 m at 2450 MHz microwave frequency).

#### 2.4. Measurement of temperature in the microwave reactor

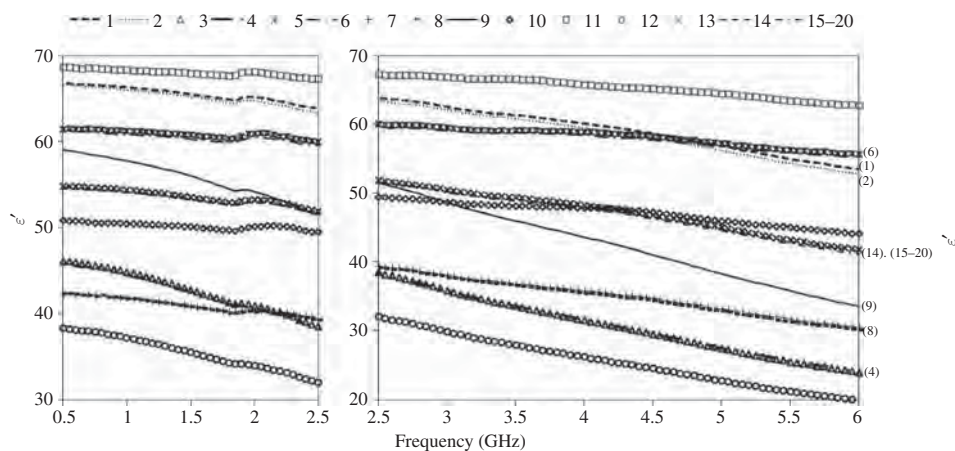
The temperature of the different solvents heated in the microwave reactor was measured to observe the rate of increase of temperature. This allowed us to observe the effect of different concentrations of solvent components (ethanol and acid) on the heating rate in terms of rate of increase of temperature in the microwave system and analyse and correlate it with the results obtained for the dielectric properties and depth of penetration. Hence, chemical combinations of 40 ml volume (0.4 ml acid + 39.6 ml ethanol solution) 0.7 mol/dm<sup>3</sup> acid with 22% v/v alcohol, 1.3 mol/dm<sup>3</sup> acid with 22% v/v alcohol, 0.7 mol/dm<sup>3</sup> acid with 58% v/v alcohol and 1.3 mol/dm<sup>3</sup> acid with 58% alcohol were prepared. These solvent combinations were heated in mono-mode microwave system (Star System 2, CEM Corporation, Matthews, NC, USA) with nominal power of 800 W at 2450 MHz frequency in a round-bottomed tubular vessel made of borosilicate glass. The actual (calibrated) power of the system is 710.5 W [20]. The experiments were conducted at 10% power level (71.05 W) for 240 s where temperature was recorded at an interval of 30 s starting from 0 s to 240 s. The temperature was recorded using a fibre optic probe and rate of increase of temperature was plotted and analysed for all solvents with respect to time.

### 3. Results and discussion

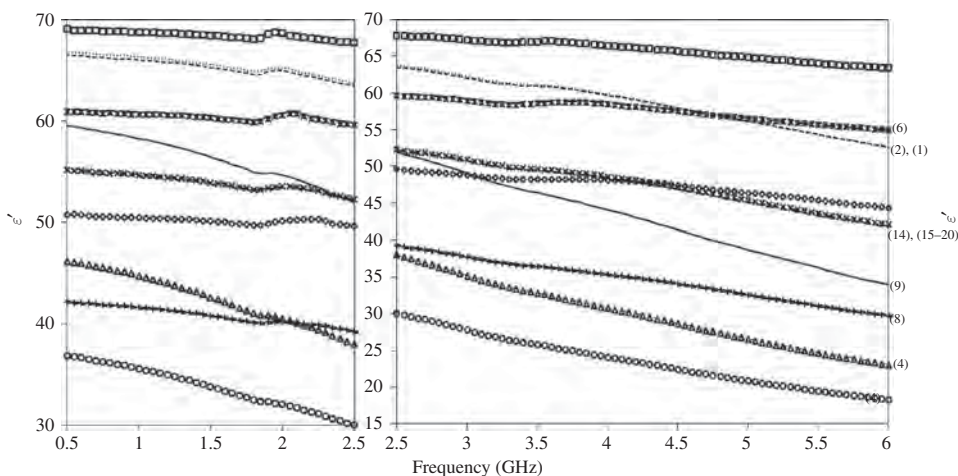
#### 3.1. Variation of the dielectric properties with microwave frequency for lower molarity HCl and citric acid (0.496 to 1.505 mol/dm<sup>3</sup>)

##### 3.1.1 Dielectric constant

Figure 1 presents the variations of  $\epsilon'$  for HCl and citric acid (for molar range 0.496–1.505 mol/dm<sup>3</sup>). The dielectric constant decreased with an increase of frequency irrespective of the acid used, its molarity, temperature level and ethanol concentration. According to the information provided in 'Agilent Basics of Measuring the Dielectric Properties of Materials Application Note', the decrease in dielectric constant with frequency occurs due to the phase lag between the dipole alignment and the electric field. From the graph, one can deduce that the frequency range chosen for the experiment is in the relaxation region as demonstrated by previous authors as well. The rate of decrease of dielectric constant was different with the different treatment combinations of factors. It can also be observed that irrespective of their higher or lower acidic nature, magnitude of  $\epsilon'$  at different frequencies was not affected by the variation of molarity for both the acids



(a)



(b)

Figure 1. Variation of  $\epsilon'$  with different combinations of (a) HCl acid and ethanol concentrations and (b) citric acid and ethanol concentrations at different temperatures.

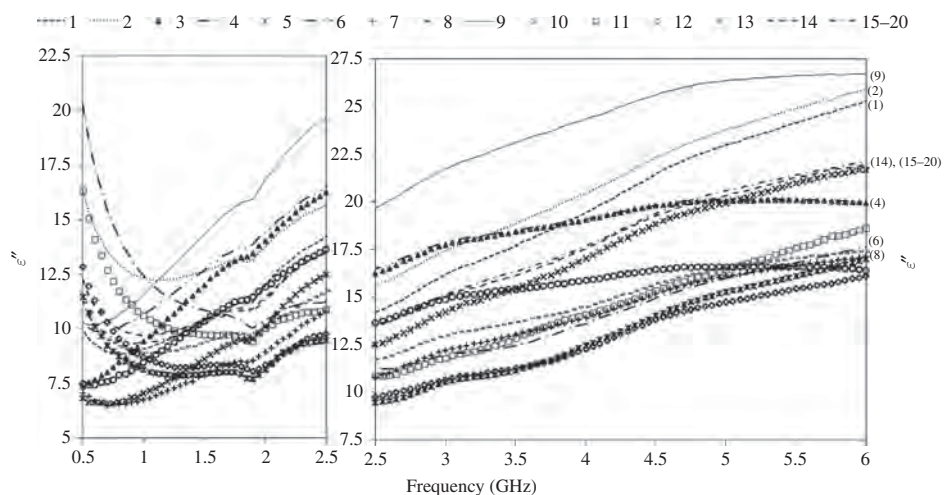
used for this range. For both acids, experiment no. 1 and 2; 3 and 4; 5 and 6; 7 and 8; 13, 14 and 15–20 have similar dielectric constant curves, which demonstrates that there is least effect of acid molarity on dielectric constant at constant temperature level and ethanol concentration for same acid type. The dielectric constant decreased with the increase in the concentration of ethanol, which was also observed at constant temperature levels. Hence, the curves for 3 and 4, and 7 and 8 are lower than the curves for 1 and 2, and 5 and 6, respectively. The figures also show that the rate of decrease of  $\epsilon'$  with increase of frequency varies with the variation of temperature level and ethanol concentration. For the range used in this study, medium to higher concentrations of ethanol and medium to lower temperature levels lead to steepest decrease in  $\epsilon'$  with increase in frequency. Hence, to obtain a higher dielectric constant for the entire range of microwave frequencies, a lower ethanol concentration and a higher temperature should be selected.



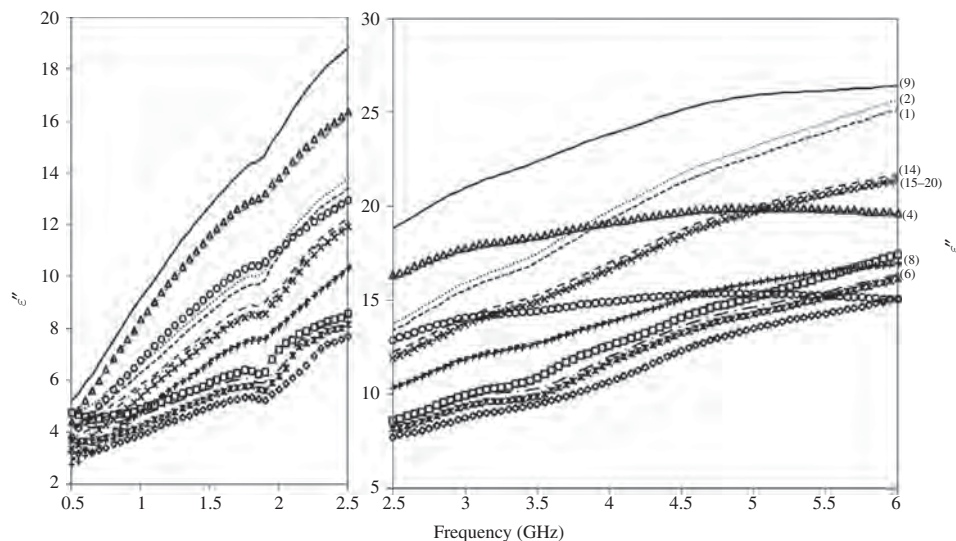
As shown in Figure 1, the  $\epsilon'$  values for citric acid and HCl are comparable and are in the same range, so the dissipation factors can be compared to obtain and interpret information about the conversion of microwave energy to thermal energy [9].

3.1.2. Dielectric loss

Figure 2 presents the variations of  $\epsilon''$  for HCl and citric acid. The variation of  $\epsilon''$  with an increase in microwave frequency for different combinations of acid and ethanol concentration and temperature levels is different for both the acids. For both acids, it can be observed that  $\epsilon''$  is higher at higher frequencies. However, for certain treatment



(a)



(b)

Figure 2. Variation of  $\epsilon''$  with respect to different combinations of (a) HCl acid and ethanol and (b) citric acid and ethanol at different temperatures.

combinations with HCl, the  $\epsilon''$  is higher at certain lower frequency levels; it decreases until 1.5 or 2 GHz and then starts increasing again. For citric acid, in the range of microwave used during the study, there is constant increase in  $\epsilon''$  with increase in microwave frequency. However, the curves for treatment combinations 5, 6 and 11 seem to be slightly decreasing with an increase in frequency, at the lowest frequency levels considered during the study, and then follow a constant increase with increase in frequency. This observation indicates multiple relaxation peaks, representative of both the presence of chemical constituents acid and alcohol, and their interaction effect on the dielectric loss [9]. From Figures 2(a) and (b), it appears that for HCl, all the factors are affecting the variation of  $\epsilon''$ , as all the curves for different combinations are different. However, for citric acid, curves for experiments 1 and 2; 3 and 4; 5 and 6; and 7 and 8 are very similar, so out of all the factors, the effect of variation of acid within the factorial range at constant temperature and ethanol level is negligible. Also, at lower frequency for citric acid, the  $\epsilon''$  for all the combinations are quite similar, which implies that at lower frequencies, the effect of the factors considered in this experiment on  $\epsilon''$  is negligible. However, for HCl, even at lower frequencies, there is a significant effect of factors. At higher frequencies, the pattern of variation of dielectric loss value for different experimental combinations using both HCl and citric acid is very similar, which might be attributed to the higher significant contribution of the other two factors maintained constant; ethanol concentration and temperature levels.

### 3.1.3. Dissipation factor

Dissipation factor is calculated based on both the dielectric constant and dielectric loss (Equation 3). The plots are, thus, based on both values. As the dielectric constant plots for both the acids are quite similar, the pattern of variation for dissipation factor with respect to frequency for the different acids is very similar to the dielectric loss plot, which are different for different acids. For HCl and ethanol combinations, the dissipation factor is initially high, which decreases with an increase in frequency up to 1 GHz and then constantly increases with increase in frequency. However, for citric acid, the dissipation factor constantly increases with increase in frequency within the frequency range used for this study. The variations in the dissipation factor caused by the independent factors are lower at lower frequencies and increase with the increase in frequency for both acids. For both acids and alcohol combinations, it can be observed that at all frequencies, a higher alcohol concentration leads to a higher dissipation factor (which can be correlated to higher heating rate), as can be observed in the case of experiments 3, 4 and 12 (Figures 3 (a) and (b)). However, at higher temperature, a lower dissipation factor was observed (experiments 10, 5 and 6; Figures 3(a) and (b)), even when a higher concentration of alcohol was used (experiments 7 and 8, Figures 3(a) and (b)). There was no significant visible effect interpreted from the plots for different molarities of acids on the dissipation factor for the range of acid molarity used during these experiments. Ethanol concentration and temperature levels, and their interaction appear to be the dominant factors in this case, so the pattern of variation of dissipation factor for the solution is governed by thermal and dielectric properties of ethanol and variation of dielectric properties of ethanol at different temperatures. For ethanol, the relaxation frequency, where the value of dielectric loss value for a substance is the highest, has been reported to vary from less than 2.45 GHz at low temperatures and to greater than 2.45 GHz at higher temperatures. Lower molecular weight polar organic molecules such as ethanol have been found to behave differently than the general thermal runaway phenomenon (a state where an increase in temperature

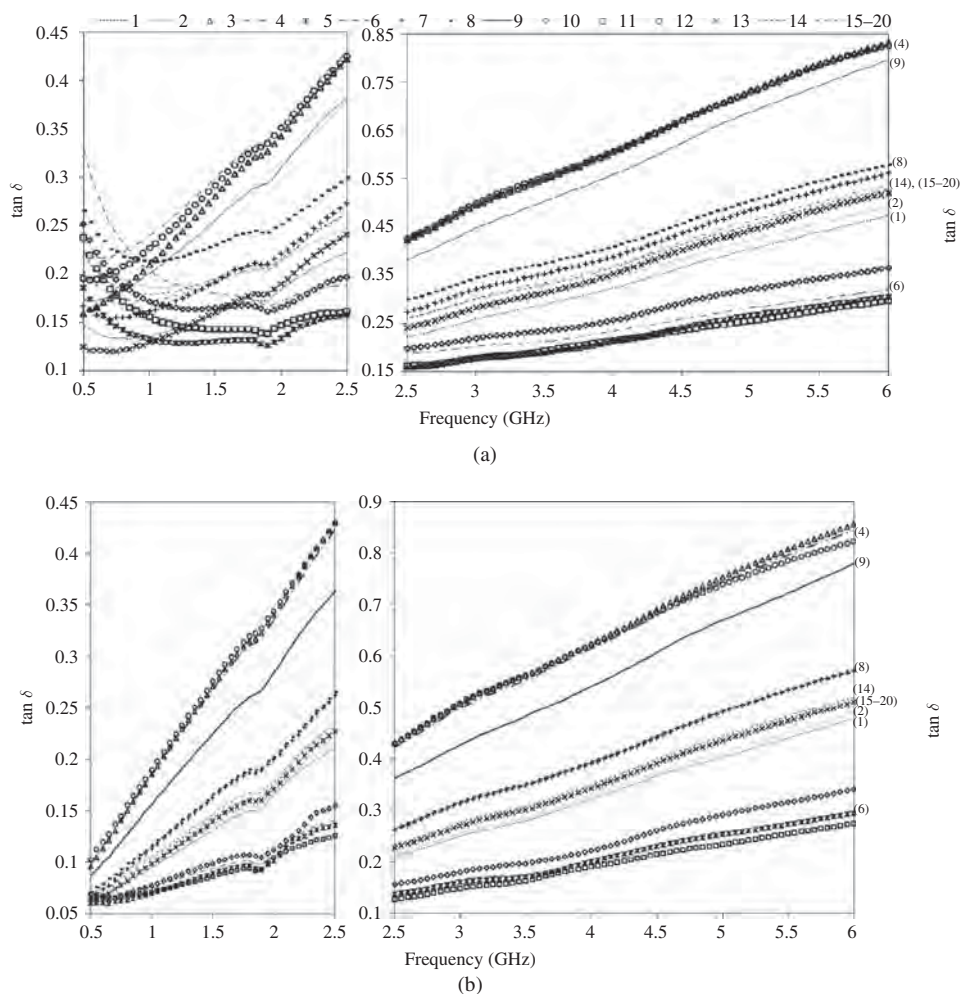


Figure 3. Variation of  $\tan \delta$  with different combinations of (a) hydrochloric acid and (b) citric acid and ethanol at different temperatures.

of a substance leads to a circumstance causing further increase in temperature). In these cases, relaxation times are very short and raising the temperature leads to a reduction in loss tangents [9]. It can be observed that temperatures such as 70.3°C or 343.45 K are pretty near to the boiling point of pure ethanol (~78.1°C or 351.25 K) and thermal runaway conditions have been found to diminish as the temperature increases and approaches towards the boiling point of the lower molecular weight polar organic solution [9], as was the case in the present study.

### 3.2. Screening analysis for dielectric constant, dielectric loss and dissipation factor for lower range of molarity of HCl and citric acids

The microwave frequency of 2450 MHz is used in domestic ovens and for extraction applications. The values of dielectric constant, dielectric loss, dissipation factor and depth of penetration for the different experimental combinations of CCDs have been summarised in Table 3 and the analysis of the screening design generated using the factorial

Table 3. Average value of dielectric constant, dielectric loss, dissipation factor and depth of penetration at various central composite design combinations of different ethanol concentrations, temperature levels and HCl and citric acid concentrations.

Expt. Comb. No.	HCl					Citric acid						
	$\epsilon'$	$\epsilon''$	$\tan \delta$	$\delta p/m$	$\epsilon'$	$\epsilon''$	$\tan \delta$	$\delta p/m$	$\epsilon'$	$\epsilon''$	$\delta p/m$	
1	63.940 ± 0.415	14.037 ± 0.315	0.2196 ± 0.0051	0.0111 ± 0.0003	63.649 ± 0.668	13.193 ± 0.446	0.2073 ± 0.0073	0.0118 ± 0.0004	63.921 ± 0.091	13.598 ± 0.156	0.2127 ± 0.0025	0.0115 ± 0.0001
2	63.469 ± 0.680	15.518 ± 0.434	0.2445 ± 0.0073	0.0101 ± 0.0003	38.296 ± 0.720	16.161 ± 0.666	0.4219 ± 0.0191	0.0076 ± 0.0003	38.300 ± 0.612	15.900 ± 0.705	0.4150 ± 0.0196	0.0077 ± 0.0003
3	38.753 ± 0.612	16.069 ± 0.341	0.4147 ± 0.0110	0.0077 ± 0.0002	59.693 ± 0.670	8.025 ± 0.382	0.1344 ± 0.0066	0.0188 ± 0.0009	60.005 ± 0.240	8.359 ± 0.115	0.1393 ± 0.0020	0.0180 ± 0.0003
4	38.644 ± 0.556	16.335 ± 0.498	0.4227 ± 0.0143	0.0076 ± 0.0002	39.417 ± 0.874	10.085 ± 0.592	0.2558 ± 0.0160	0.0122 ± 0.0007	39.389 ± 0.800	10.200 ± 0.578	0.2589 ± 0.0156	0.0121 ± 0.0007
5	60.079 ± 0.194	9.416 ± 0.473	0.1567 ± 0.0079	0.0161 ± 0.0008	52.248 ± 0.499	18.591 ± 1.070	0.3558 ± 0.0208	0.0077 ± 0.0004	49.699 ± 0.272	7.577 ± 0.113	0.1524 ± 0.0024	0.0181 ± 0.0003
6	59.595 ± 0.918	11.198 ± 0.396	0.1879 ± 0.0073	0.0135 ± 0.0005	67.825 ± 0.195	8.428 ± 0.235	0.1243 ± 0.0035	0.0190 ± 0.0005	30.265 ± 0.846	12.769 ± 0.661	0.4218 ± 0.0248	0.0086 ± 0.0005
7	39.670 ± 0.499	10.638 ± 0.478	0.2682 ± 0.0125	0.0116 ± 0.0005	52.471 ± 0.588	11.718 ± 0.540	0.2233 ± 0.0106	0.0121 ± 0.0006	52.405 ± 0.642	12.060 ± 0.432	0.2301 ± 0.0087	0.0117 ± 0.0004
8	39.389 ± 0.667	11.619 ± 0.492	0.2951 ± 0.0135	0.0106 ± 0.0005	52.297 ± 0.279	11.868 ± 0.175	0.2269 ± 0.0036	0.0119 ± 0.0002	52.297 ± 0.279	11.868 ± 0.175	0.2269 ± 0.0036	0.0119 ± 0.0002
9	51.716 ± 0.474	19.473 ± 0.264	0.3765 ± 0.0062	0.0073 ± 0.0001								
10	49.549 ± 0.853	9.692 ± 0.706	0.1955 ± 0.0146	0.0142 ± 0.0010								
11	67.357 ± 0.251	10.793 ± 0.223	0.1602 ± 0.0034	0.0148 ± 0.0003								
12	32.245 ± 0.183	13.479 ± 0.106	0.4180 ± 0.0041	0.0084 ± 0.0001								
13	52.072 ± 0.362	12.338 ± 0.354	0.2369 ± 0.0070	0.0114 ± 0.0003								
14	51.756 ± 0.655	13.370 ± 0.720	0.2584 ± 0.0143	0.0106 ± 0.0006								
15-20	51.727 ± 0.251	13.399 ± 0.182	0.2590 ± 0.0037	0.0105 ± 0.0001								

Notes: For all the cases reported above, the reported uncertainty is based on the standard uncertainty multiplied by a coverage factor  $k = 2$ , providing a level of confidence of approximately 95%.  $\epsilon'$  = Dielectric constant,  $\epsilon''$  = Dielectric loss,  $\tan \delta$  = Dissipation factor,  $\delta p$  = Depth of penetration (m).

parts of CCDs with ethanol, temperature and the acids (HCl and citric acid) at 2450 MHz was done using the JMP 8 software. The analysis was done to distinguish the effect of linear, bilinear and interactive effects of three and four factors. However, during the analyses, the linear and interactive effects with individual  $p$ -value  $<0.05$  were detected and used to generate the predictive model with the significant linear and interactive terms. The ANOVA analysis of the predictive model and corresponding terms and their consequent parameter estimates at coded factor levels  $(-1, 1)$  have been reported in Table 4.

Table 4. ANOVA for the screening analyses with different ethanol concentrations, temperature levels and HCl and citric acid molarity levels and the corresponding parameter estimates at coded factor levels  $(-1, 1)$  for the predictive model.

Source of variation	$df$	SS	MS	$F$ ratio	Prob $> F$	Estimate
<b>For <math>\epsilon'</math> (<math>R^2 = 0.9984</math>, Adj <math>R^2 = 0.9983</math>, RMSE = 0.4863)</b>						
Intercept						50.388
EtOH	1	6244.52	6244.52	26399.94	$<0.0001$	-11.406
Temp	1	25.82	25.82	109.18	$<0.0001$	-0.734
EtOH*Temp	1	71.15	71.15	300.82	$<0.0001$	1.218
Model	4	6342.17	1585.54	6703.20	$<0.0001$	
Error	43	10.17	0.24			
LOF	11	0.95	0.09	0.30	0.9813	
Pure error	32	9.23	0.29			
Total	47	6352.35				
<b>For <math>\epsilon''</math> (<math>R^2 = 0.9846</math>, Adj <math>R^2 = 0.9809</math>, RMSE = 0.3998)</b>						
Intercept						12.522
Temp	1	319.37	319.37	1998.24	$<0.0001$	-2.579
EtOH	1	35.01	35.01	219.04	$<0.0001$	0.854
Acid [HCl]	1	16.25	16.25	101.68	$<0.0001$	0.582
Acid level	1	4.88	4.88	30.52	$<0.0001$	0.319
EtOH*Temp	1	1.24	1.24	7.77	0.0082	-0.161
Temp*Acid [HCl]	1	1.79	1.79	11.23	0.0018	0.193
EtOH*Acid [HCl]	1	4.10	4.10	25.66	$<0.0001$	-0.292
EtOH*Acid level	1	1.58	1.58	9.87	0.0033	-0.181
Acid [HCl]*Acid level	1	2.88	2.88	17.99	0.0001	0.245
Model	9	387.1	43.01	269.11	$<0.0001$	
Error	38	6.07	0.16			
LOF	6	0.81	0.13	0.82	0.5640	
Pure error	32	5.27	0.16			
Total	47	393.17				
<b>For <math>\tan \delta</math> (<math>R^2 = 0.9946</math>, Adj <math>R^2 = 0.9931</math>, RMSE = 0.0083)</b>						
Intercept						0.2659
EtOH	1	0.2929	0.2929	4287.32	$<0.0001$	0.0781
Temp	1	0.1394	0.1394	2039.89	$<0.0001$	-0.0539
Acid [HCl]	1	0.0051	0.0051	73.91	$<0.0001$	0.0103
Acid level	1	0.0018	0.0018	26.11	$<0.0001$	0.0061
EtOH*Temp	1	0.0205	0.0205	300.28	$<0.0001$	-0.0207
EtOH*Acid [HCl]	1	0.0008	0.0008	11.92	0.0014	-0.0041
Temp*Acid [HCl]	1	0.0011	0.0011	15.43	0.0004	0.0047
Acid [HCl]*Acid level	1	0.0013	0.0013	19.55	$<0.0001$	0.0053

(continued)



Table 4. (Continued).

Source of variation	df	SS	MS	F ratio	Prob > F	Estimate
Model	10	0.4633	0.0463	678.11	<0.0001	
Error	37	0.0025	0.0001			
LOF	5	0.0002	0.00005	0.68	0.6395	
Pure error	32	0.0023	0.00007			
Total	47	0.4659				
<b>For <math>\delta p</math> (m) (<math>R^2 = 0.9889</math>, Adj <math>R^2 = 0.9855</math>, RMSE = 0.0004)</b>						
Intercept						0.0117
Temp	1	0.0003	0.0003	1569.40	<0.0001	0.0024
EtOH	1	0.0002	0.0002	1252.30	<0.0001	-0.0021
Acid [HCl]	1	0.00003	0.00003	146.44	<0.0001	-0.0007
Acid level	1	0.000007	0.000007	37.97	<0.0001	-0.0004
Temp*EtOH	1	0.000007	0.000007	38.20	<0.0001	-0.0004
Temp*Acid [HCl]	1	0.00001	0.00001	55.89	<0.0001	-0.0005
EtOH*Acid [HCl]	1	0.00001	0.00001	58.08	<0.0001	0.0005
Temp*Acid level	1	0.000002	0.000002	10.14	0.003	-0.0002
EtOH*Acid level	1	0.000002	0.000002	14.09	0.0006	0.0002
Acid [HCl]*Acid level	1	0.000003	0.000003	15.45	0.0004	-0.0002
Temp*EtOH*Acid [HCl]	1	0.000002	0.000002	11.36	0.0018	0.0002
Model	11	0.00055	0.00005	291.76	<0.0001	
Error	36	0.000006	1.706e-7			
LOF	4	1.169e-6	2.92e-7	1.88	0.1379	
Pure error	32	4.97e-6	1.55e-7			
Total	47	0.00055				

Notes: Experiment was designed considering Temperature in °C and acid concentration in molarity and not molality.  $df$  = degree of freedom, SS = Sum of squares, MS = Mean square, Adj  $R^2$  = Adjusted  $R^2$ , RMSE = Root mean square error, EtOH = Ethanol concentration (% v/v), Temp = Temperature levels (°C), Acid level = Acid concentration level (M, mol/dm<sup>3</sup>),  $\epsilon'$  = Dielectric constant,  $\epsilon''$  = Dielectric loss,  $\tan \delta$  = Dissipation factor,  $\delta p$  = Depth of penetration (m).

It can be observed that ethanol, temperature and the bilinear interactive term of ethanol and temperature are the significant terms contributing to the predictive model for the dielectric constant. Overall, the model was significant at the  $p$ -value 0.05 and had an  $R^2$  value of 0.9984 with insignificant lack of fit (value of Prob >  $F$  was 0.9813). It can be interpreted that by increasing the concentration levels of ethanol and the temperature from -1 to 1 coded level, the value of dielectric constant can be greatly decreased, mainly due to the ethanol, which has the highest negative value of regression coefficient (parameter estimate) (Table 4). From the values of parameter estimates or regression coefficients, it can be observed that the most influential factor is ethanol with greatest negative value of regression coefficient, -11.406, followed by the interactive term of ethanol and temperature with 1.218 regression coefficient, and lastly, temperature with -0.734 regression coefficient. The contribution of acids to the value of dielectric constant was not significant, which was also observed from Figures 1(a) and (b). Hence, it can be concluded that ethanol and temperature are the only contributing factors governing the dielectric constant. Also, it can be noted that the increase of ethanol concentration and temperature leads to a decrease of dielectric constant values at 2450 MHz microwave frequency.

At the microwave frequency of 2450 MHz, there were many factors that were contributing to the final dielectric loss value other than ethanol concentration and

temperature levels (Table 4). Acid type and acid level were also found to be significant factors. The predictive model generated was significant at  $p$ -value 0.05 with  $R^2$  value 0.9846 with insignificant lack of fit (value of  $\text{prob} > F$  was 0.564). Ethanol\*temperature, temperature\*acid type, ethanol\*acid type, ethanol\*acid level, acid type\*acid level were the interactive terms contributing to the predictive model. From the parameter estimates it can be interpreted that, temperature is the most influential factor, followed by ethanol concentration, acid type and acid level. From Table 4 it can be observed that, for this range of molarity (0.496 to 1.505 mol/dm<sup>3</sup>) among HCl and citric acid, HCl has a positive effect on dielectric loss, i.e. dielectric loss values are higher in the case of HCl at lower ethanol concentrations than the cases with citric acid, as observed from Table 3 and analysis of screening design. It can be attributed to the polar nature of HCl which is higher than that of citric acid and also to the higher acidic properties. It can be observed that with an increase in temperature, the dielectric loss decreases for the experimental combinations with the same composition (same concentration of ethanol and acid). For example, in Figure 2, it can be observed that the curves for experimental combinations 1 and 2, and 9 show higher values of dielectric loss than curves for treatment combinations 3 and 4, and 10, respectively. Among the interactive factors, considering the parameter estimate, it can be interpreted that, the interaction between ethanol concentration and acid, and acid and acid molarity are the most significant contributing factors. While the individual or linear effect of ethanol concentration and acid concentration level is positive, the interactive effect of ethanol with acid type HCl, and ethanol with acid molarity level, is negative. It implies that, while dielectric loss increases with increase in ethanol concentration, the increase is lower with acid, and with higher acid molarity level. The increase in ethanol concentration with higher molarity level of acid leads to a decrease in dielectric loss of the resulting solution combination. Higher ethanol concentration with higher temperature also leads to a lowering of dielectric loss. However, interaction of higher temperature with acid type HCl at higher acid molarity can lead to an increase in the dielectric loss. The combination of all these significant terms explains the diversity of the dielectric loss values for various combinations of acid and ethanol concentration and temperature levels for treatment combinations with HCl (Figure 2(a)) and citric acid (Figure 2(b)). It also explains the multiple relaxation peaks for the alcohol and acid combination. As discussed in the previous section, the effect of citric acid on the variation of dielectric loss is negligible at lower frequencies upto 2450 MHz as well. Out of the two acids, HCl is the significant acid type, due to its more acidic nature; leading to the formation of ions which lead to ionic relaxation predominant at lower frequencies, which contributes to the dielectric loss.

At 2450 MHz frequency, the dissipation factor is influenced by all the factors, ethanol concentration, temperature levels, acid type and acid molarity levels. The predictive model was significant at  $p$ -value 0.05 with  $R^2$  value of 0.9936 and insignificant lack of fit (value of  $\text{prob} > F$  is 0.2360). Based on parameter estimates (Table 4), it can be interpreted that, with an increase in ethanol concentration and increase in acid molarity for HCl, the dissipation factor increases. However, with the increase in temperature of the solvent combination, the dissipation factor decreases. The pattern of effects of these factors is very similar to the case of dielectric loss. The influential interactive effects include interaction of ethanol concentration with temperature, ethanol concentration with type of acid, temperature levels with acid type and acid type with its molarity levels. The most influential interactive effect is the interactive effect of ethanol concentration with different temperature levels. The negative value of the parameter estimate for interactive effect implies that when solvents with higher concentration of ethanol were heated to

higher temperature, it lead to decrease of dielectric loss as was mentioned in the previous section on dissipation factor, which implies that the rate of increase of temperature of combinations with higher ethanol concentration decreases at higher temperature. Also, combination of higher ethanol concentration with HCl acid had a negative influence over the dielectric loss, whereas combination of higher temperature levels and acid type HCl, and acid type HCl with its higher molarity had positive influence over the dielectric loss. The more significant effect of HCl might be explained by the higher acidic nature of HCl, which leads to the fluidity of ions leading to a greater ionic contribution to the dielectric loss. In Figure 3(a), it can be observed that apart from the highly visible effect of the higher concentration of ethanol and lower temperature level, there is diversity in the resulting dissipation factor due to the different treatment combinations and their different significant interactive effects.

At 2450 MHz, it can be observed that the depth of penetration is affected by all the factors including temperature level, alcohol concentration, acid type and acid level and also their interactive effects (Table 4). According to the observed values of parameter estimates, temperature level is the most influential factor determining the depth of penetration followed by ethanol concentration, acid type and acid concentration level (Table 4). Among the interactive effects, the most influential combinations of parameters are temperature with acid type and ethanol with acid type, followed by temperature level with ethanol. The least significant interactive effects include combinations of factors temperature with acid level, ethanol with acid level, acid type with acid level and temperature with ethanol and acid type; all of which appear to be equally significant. Depending on the sign and values of the parameter estimate, it can be observed that the depth of penetration increases with increase in temperature, which implies that the microwave heating rate decreases with increase in temperature which agrees with the observation in terms of dissipation factor. The same results can be observed in Table 3. With increased ethanol concentration and acid concentration levels, the depth of penetration decreases, which implies that microwave heating rate increases. This agrees with the previous discussions and results mentioned in Table 3 also. As interpreted from the negative value of parameter estimate for acid type observed from the values in Table 4, the values of dissipation factor for HCl are lower (or in some cases similar) than citric acid at the same acid concentration level, ethanol concentration and temperature level. Hence, the heating rate is expected to be comparatively higher for HCl than for citric acid. Similarly, it can be observed that the combined effect of temperature with ethanol concentration, temperature level with acid type HCl, temperature with acid level, and acid type HCl with acid level is negative, leading to an increase in heating rate; whereas the combined effect of ethanol with HCl acid, ethanol with acid level and temperature with ethanol and HCl acid leads to a decrease in heating rate. The interactive effect of factors can be observed at higher levels of ethanol concentration and higher temperature levels in Table 3.

### **3.3. Central composite design with response surface methodology analysis for dissipation factor in lower range of molarity of HCl and citric acids**

The aim of the CCD-RSM analysis of the dielectric properties of the solvents was the interpretation of the heating rate of the solvents through the dissipation factor. For other response parameters including dielectric constant, dielectric loss and depth of penetration, lack of fit for the model obtained through response surface analysis was found to be significant; hence has not been reported in this manuscript. A CCD-RSM allows the analysis of scientific data with the minimum number of treatment combinations possible [21]. As mentioned before, the values of dielectric constant, dielectric loss, dissipation

factor and depth of penetration for the different experimental combinations of CCDs have been summarised in Table 3. In the previous section, it was interpreted that a higher heating rate can be observed for HCl than for citric acid, possibly because HCl is a stronger acid than citric acid. However, it was found that the dissipation factor for both the acids at 2.45 GHz was greater than 0.1, which is the dissipation factor of water at this frequency. It has been mentioned by Gabriel et al. [9], that as water is a useful dielectric heating medium and is the basis of the successful energy transfer of microwaves in the food industry, any solvent having a loss tangent value greater than 0.1 is an appropriate solvent in dielectric heating [9]. Hence, for the interpretation of the heating rate of combinations of individual acids with ethanol in terms of dissipation factor, an analysis of the CCD using RSM was applied separately for citric acid and HCl. The parameters estimate or regression coefficients for the predictive model were deduced and the corresponding ANOVA analysis has been summarised in Table 5.

In case of HCl, as observed in Table 5, influential parameters included temperature levels, ethanol concentrations and HCl molarity. Overall, a predictive model with

Table 5. ANOVA for the CCD-RSM analyses of dissipation factor with respect to different ethanol concentrations, temperature levels and HCl and citric acid concentrations and the corresponding parameter estimates at coded factor levels for the predictive model.

Source of variation	<i>df</i>	SS	MS	<i>F</i> ratio	Prob > <i>F</i>	Estimate
<b>For combinations with HCl (<math>R^2 = 0.9910</math>, Adj <math>R^2 = 0.9894</math>, RMSE = 0.0081)</b>						
Intercept						0.2590
Temp	1	0.1070	0.1070	1639.57	<0.0001	-0.0511
EtOH	1	0.2310	0.2310	3538.70	<0.0001	0.0751
HCl	1	0.0036	0.0036	54.30	<0.0001	0.0093
Temp*EtOH	1	0.0090	0.0090	137.50	<0.0001	-0.0193
Temp*Temp	1	0.0041	0.0041	63.13	<0.0001	0.0098
EtOH*EtOH	1	0.0051	0.0051	78.04	<0.0001	0.0109
HCl*HCl	1	0.0006	0.0006	9.56	0.0032	-0.0038
Model	9	0.3608	0.0401	614.05	<0.0001	
Error	50	0.0033	0.0001			
Lack of Fit	5	0.0006	0.0001	2.0234	0.0934	
Pure error	45	0.0027	0.0001			
Total	59	0.3641				
<b>For combinations with citric acid (<math>R^2 = 0.9942</math>, Adj <math>R^2 = 0.9932</math>, RMSE = 0.0098)</b>						
Intercept						1.2546
Temp	1	0.2507	0.2507	2598.17	<0.0001	-0.0782
EtOH	1	0.5113	0.5113	5299.49	<0.0001	0.1117
Temp*EtOH	1	0.0291	0.0291	301.38	<0.0001	-0.0348
Temp*Temp	1	0.0101	0.0101	104.30	<0.0001	0.0153
EtOH*EtOH	1	0.0309	0.0309	320.31	<0.0001	0.0267
Model	9	0.8296	0.0922	955.44	<0.0001	
Error	50	0.0048	0.0001			
Lack of fit	5	0.0009	0.0002	2.13	0.0796	
Pure error	45	0.0039	0.0001			
Total	59	0.8345				

Notes: Experiment was designed considering Temperature in °C and acid concentration in molarity and not molality. *df* = degree of freedom, SS = Sum of squares, MS = Mean square, Adj  $R^2$  = Adjusted  $R^2$ , RMSE = Root mean square error, EtOH = Ethanol concentration (% v/v), Temp = Temperature levels (°C), Acid level = Acid concentration level (M).

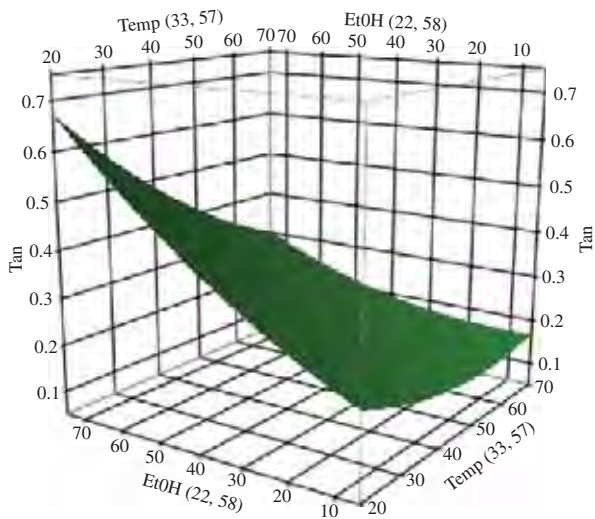
linear, quadratic and bilinear terms was found to be significant at  $\alpha < 0.05$ , with  $R^2$  value 0.9910 and insignificant lack of fit ( $p$ -value was 0.0934). High  $R^2$  value with insignificant lack of fit indicates the reliability and adequacy of the predictive model to represent the system. The predictive model for HCl in terms of coded factor levels was

$$\begin{aligned} \tan \delta = & 0.2590 - 0.0511 \times \text{Temp} + 0.0751 \times \text{EtOH} + 0.0093 \times \text{HCl} - 0.0193 \\ & \times \text{Temp} \times \text{EtOH} + 0.0098 \times \text{Temp}^2 + 0.0109 \times \text{EtOH}^2 - 0.0038 \times \text{HCl}^2 \end{aligned} \quad (5)$$

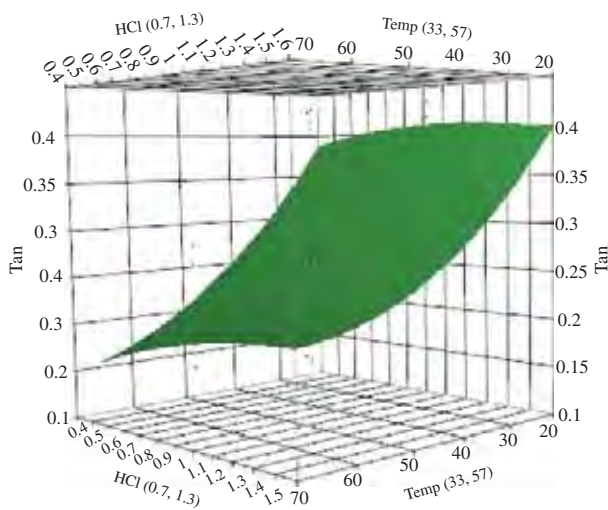
Response surfaces were plotted representing the effect of ethanol concentration and temperature levels, and effect of HCl molarity and temperature levels on the dissipation factor. Among the linear effects, temperature level effect was negative, which indicated that with an increase in temperature, the dissipation factor will decrease (Figures 4(a) and (b)), which is a similar observation as discussed in the previous section on dissipation factor. The parameter estimates for ethanol concentrations (numerical value (excluding the positive or negative sign) of parameter estimate was higher than that of temperature) and HCl molarity (numerical value of parameter estimate was lower than that of temperature) were found to be positive, which implied that with an increase in the concentration of ethanol (Figure 4(a)) and HCl molarity, the dissipation factor will increase, where ethanol concentration can be interpreted to be more influential than the HCl molarity. The parameter estimates for quadratic terms Temperature\*Temperature, Ethanol\*Ethanol and HCl\*HCl were all found to be significant at  $\alpha < 0.05$ , where ethanol\*ethanol was found to be most influential, followed by temperature\*temperature and then lastly HCl\*HCl, which has negative parameter estimate value. The positive values for the parameter estimates for temperature\*temperature and ethanol\*ethanol imply that there is a point of minimum (response decreases until this point and then increases) on the response surface of dissipation factor plotted in terms of these two factors. However, based on the negative parameter estimate value for quadratic HCl term, it can be interpreted that there is a point of maximum for dissipation factor plotted in terms of HCl concentration, which implies that the dissipation factor increases with increase in HCl molarity but then it decreases with further increase in HCl molarity. Only one bilinear term temperature\*ethanol is significant, whose parameter estimate has a negative value, which implies that at higher values of both ethanol concentration and temperature, the resulting dissipation factor will have lower value which can also be seen in Figures 3(a) and 4(a), and has been discussed before in the previous section on dissipation factor. For lower molecular weight compound solutions, the increase in temperature leads to a lowering of the dissipation factor [9]. The values of dissipation factor for treatment combinations with higher concentration of ethanol and higher temperatures (treatment combinations 7, 8 and 5, 6) are lower than the combinations with lower ethanol concentrations and temperatures (treatment combinations 3, 4 and 1, 2, respectively).

In case of citric acid, when CCD was analysed with RSM, an exponential transformation was applied to increase the insignificance of lack of fit. Temperature and ethanol concentrations were detected as the only contributing factors for the determination of dissipation factor of ethanol and citric acid combinations. Overall, the predictive model with linear, quadratic and bilinear terms was found to be significant at  $\alpha < 0.05$  with  $R^2$  0.9942 and insignificant lack of fit ( $p$ -value was 0.0796). The predicted model with coded levels of factors was





(a)



(b)

Figure 4. Response surface plot representing effect of (a) ethanol concentration (% v/v, EtOH) and temperature levels (°C, Temp) (b) temperature levels (°C, Temp) and acid molarity (M, HCl) on dissipation factor (Tan) for central composite design with hydrochloric acid.

$$\begin{aligned} \text{Exp}(\tan \delta) = & 1.2546 - 0.0782 \times \text{Temp} + 0.1117 \times \text{EtOH} - 0.0348 \times \text{Temp} \\ & \times \text{EtOH} + 0.0153 \times \text{Temp}^2 + 0.0267 \times \text{EtOH}^2 \end{aligned} \quad (6)$$

Based on parameter estimate values, it can be interpreted that, the linear terms temperatures and ethanol concentration, bilinear term temperature\*ethanol and quadratic terms

temperature\*temperature and ethanol\*ethanol are influential terms at  $\alpha < 0.05$ . Ethanol was interpreted to be a more influential factor than temperature, and it can also be interpreted that the increase of ethanol concentration (Figure 3(b), curve for treatment combinations 3 and 4 is higher than 1 and 2) and lower temperatures (Figure 3(b), curve for treatment combinations 3 and 4 is higher than 7 and 8, respectively) can give a higher dissipation factor, leading to a higher heating rate. The response surface developed to analyse the effect of ethanol and temperature also supports the same fact (Figure 5). However, as implied by the negative value of ethanol concentration\*temperature level also observed in Figure 5, it can be interpreted that at higher temperatures, it becomes more dominant over the ethanol concentration, so even when the ethanol concentration is higher, at higher temperature, it leads to lower dissipation factor (Figure 3(b), curve for treatment combinations 3 and 4 is higher than 7 and 8, respectively) which is similar to the case of HCl as discussed before. Among the quadratic factors, ethanol concentration\*ethanol concentration is a more influential factor than temperature level\*temperature level. The positive value of parameter estimates for temperature level\*temperature level and the negative value of parameter estimate for temperature implies that there is a point of minimum for the temperature in this experimental parameters range, where the lowest value possible for the dissipation factor is present (Figure 5). During this analysis, citric acid was not found to be an influential factor.

It can be observed that during the factorial screening design analysis, among the two acids, citric acid and HCl, HCl was interpreted to result in a higher heating rate than citric acid, based on the analysis of observations of dissipation factor. Also, when the individual case was observed for the acids separately using RSM, the predictive model for HCl indicated its significant influence and citric acid was not found to be significantly contributing to the dissipation factor of the solvent combination, which is in agreement with the previous observations, according to which HCl was expected to be an influential

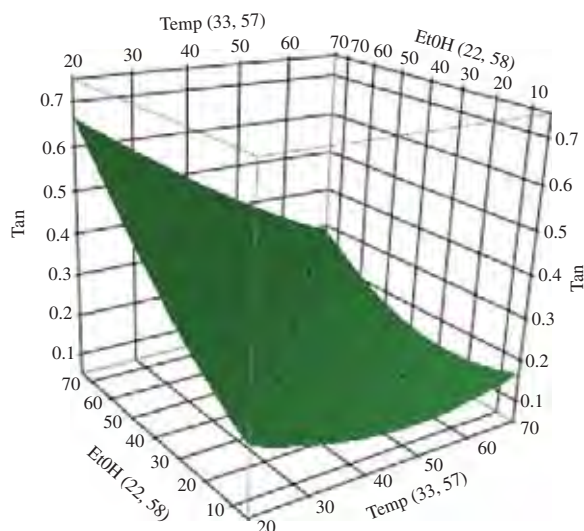


Figure 5. Response surface plot representing the effect of ethanol concentration (% v/v, EtOH) and temperature levels ( $^{\circ}\text{C}$ , Temp) on dissipation factor (Tan) for central composite design with citric acid.

acid. It seems that for weak acids, analysis with a lower range of molarity can provide sufficient information regarding their dielectric properties and there is no interaction with other factors such as ethanol concentration and temperature levels. However, for stronger acids, dielectric studies of solvent combinations with wider molarity range might be more helpful and clear. To further study the effects of different concentrations of HCl with different ethanol concentrations on dielectric properties, experiments in a broader range covering higher concentration of HCl, compared with other acid concentrations at similar molarities are highly recommended.

### 3.4. *Variation of temperature with time of application of microwave in microwave reactor*

From the dielectric measurements analysis, temperature level and ethanol concentration are the most significant factors when ethanol is combined with lower concentrations of acids. The different combinations of ethanol and acids (as mentioned in the section Measurement of Temperature in the Microwave Reactor) were heated in the microwave reactor to assess the corresponding change in temperature with time of application of microwave, which was plotted in Figure 6. The highest temperature achieved was 90°C which also prevented boiling of solutions, loss of the solvent and decrease of solvent volume. The effect of higher temperature, which was predicted to lead to a decrease of heating rate (based on dielectric properties and dissipation factor) at higher temperature was not found to significantly affect the rate of increase of temperature with time of application of microwave. This might be explained by the fact that, the temperature of 90°C or 363.15 K might not be high enough to observe any significantly visible effect of temperature on the heating rate of the solution. However, the temperature levels achieved at the same time of application of microwave on solvents with higher concentrations of ethanol were higher than those for lower concentrations of ethanol, which coincides with the dielectric analyses which implied a higher rate of heating for higher concentration of ethanol. From Figure 6, it can be observed that the separation between the plot for higher concentration of ethanol and lower concentration of ethanol increased with increase in temperature for both the acids, which implies the significant interactive effect of temperature with ethanol concentration, as discussed in the previous sections. The temperature levels achieved in case of both the acids were similar, which implies insignificant effect of lower concentrations of acids on the heating properties of the solvent combinations. Also, from the plot, it can be interpreted that the effect of the molarity level of the acids on the heating rate is also negligible at lower temperature levels and low acid molarity levels. These observations agree with the previous observations and analyses; hence proving that dielectric studies can be used as a tool for prediction of microwave heating behaviour of solvents.

## 4. Conclusion

This experimental study indicates that in case of dielectric heating of alcohol–acid combinations, interpretation of heating rate in terms of dielectric studies coincides with the reaction of the solution in a microwave reactor. However, according to our knowledge, this is one of the first studies which has confirmed that the dielectric behaviour of the solvents with combination of different constituent compounds can be used for the prediction of microwave heating behaviour. More studies will aid in a generalisation of the solvent behaviour, i.e. for this case, interaction of ethanol with different types of acids in a microwave environment. Alcohol concentration and temperature level have been

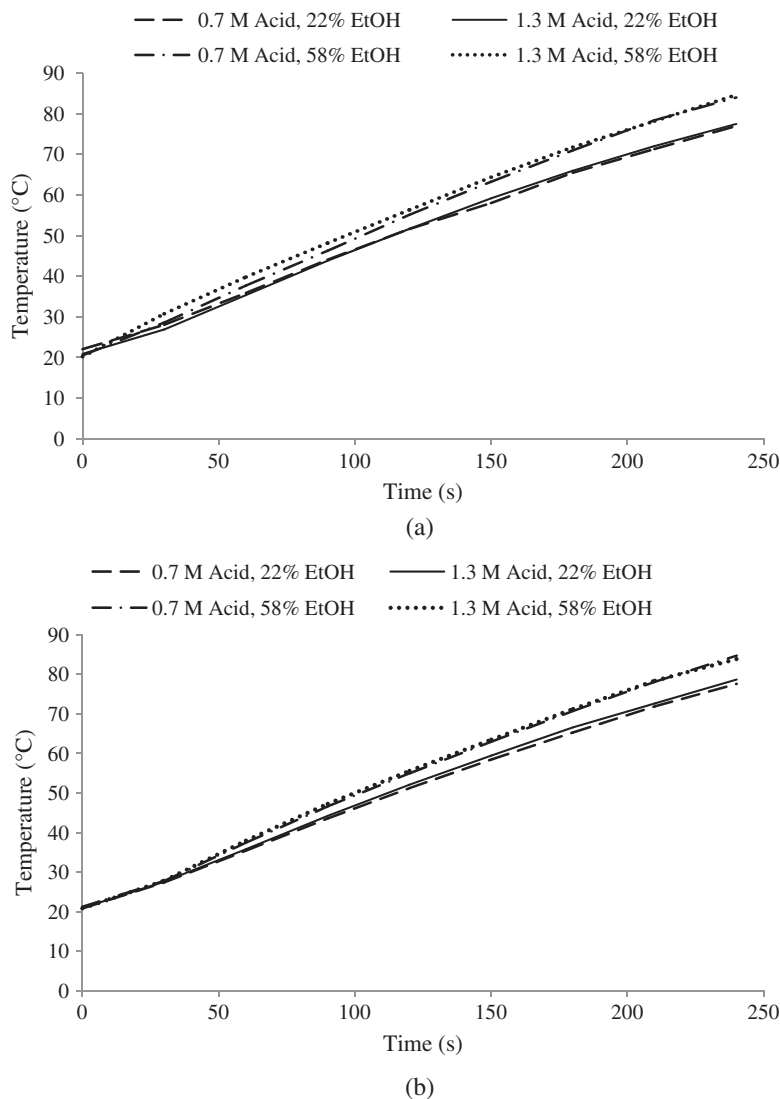


Figure 6. Plot representing effect of ethanol concentration (% v/v) and acid concentration (M) on the temperature of the solution with respect to time of application of microwave in the microwave reactor. (a) is for the combinations with citric acid and ethanol and (b) is for the combinations with HCl and ethanol. Series 1 represents the combination with 0.7 mol/dm<sup>3</sup> acid and 22% ethanol; Series 2 represents solution with 1.3 mol/dm<sup>3</sup> acid and 22% ethanol; Series 3 represents solution with 0.7 mol/dm<sup>3</sup> acid and 58% ethanol; Series 4 represents solution with 1.3 mol/dm<sup>3</sup> acid and 58% ethanol.

found to be the most significant factors affecting the dielectric properties as observed in the graphs plotted with respect to different frequencies and screening studies. Dissipation factor was also analysed using CCD-RSM for the range of different factors used in this study; however, regression analysis using RSM could not be done for other responses (dielectric constant, dielectric loss, depth of penetration). This could be explained by the unsuitability of the range of different factors used during this study for the response

surface analyses for other responses. Also in case of citric acid, at lower molarity level, acid concentration was found to be an insignificant factor for the response surface analysis and generation of predictive model for dissipation factor, which implies its insignificant role in the heating rate of ethanol and citric acid combinations at lower concentration levels of acid. However, it is highly recommended to study the dielectric properties of solvent combinations with HCl at higher acid concentration levels with a larger range of variation of acid molarity to further understand the interaction of ethanol and citric in a microwave environment. Additional studies in this area will further solidify the idea of using dielectric studies as a potential tool of interpretation of behaviour of new solvents with combination of constituents in a microwave environment, which can influence the decision regarding the use of solvents for various novel analytical studies.

### Acknowledgements

We thank Dr. G.S.V. Raghavan (James McGill Professor, Bioresource Engineering Department, Macdonald Campus, McGill University, Canada) for providing us the required equipments and facilities for the experiments. We also thank Mr. Yvan Garipey (Research Associate, Bioresource Engineering Department, Macdonald Campus, McGill University, Canada) for his expertise and support in conducting the experiments.

### References

- [1] Venkatesh MS, Raghavan GSV. An overview of microwave processing and dielectric properties of agri-food materials. *Biosys Eng.* 2004;88:1–18.
- [2] Casazza AA, Aliakbarian B, Mantegna S, Cravotto G, Perego P. Extraction of phenolics from *Vitis vinifera* wastes using non-conventional techniques. *J Food Eng.* 2010;100:50–55.
- [3] Vian MA, Fabiano-Tixier AS, Elmaataoui M, Dangles O, Chemat F. A remarkable influence of microwave extraction: enhancement of antioxidant activity of extracted onion varieties. *Food Chem.* 2011;127:1472–1480.
- [4] Jain T, Jain V, Pandey R, Vyas A, Shukla SS. Microwave assisted extraction for phytoconstituents: an overview. *Asian J Res Chem.* 2009;2:19–25.
- [5] Routray W, Orsat V. Microwave-assisted extraction of flavonoids: a review. *Food Biopro. Tech.* 2012;5(2):409–424.
- [6] Chen HC, Camire ME. Recovery of anthocyanins, pectin, and dietary fiber from cull lowbush blueberries. *J Food Qual.* 1997;20:199–209.
- [7] Karaaslan M, Ozden M, Vardin H, Turkoglu H. Phenolic fortification of yogurt using grape and callus extracts. *LWT – Food Sci Tech.* 2011;44:1065–1072.
- [8] Barnes JS, Nguyen HP, Shen S, Schug KA. General method for extraction of blueberry anthocyanins and identification using high performance liquid chromatography-electrospray ionization-ion trap-time of flight-mass spectrometry. *J Chromatogr. A* 2009;1216:4728–4735.
- [9] Gabriel C, Gabriel S, Grant EH, Halstead BSJ, Mingos DMP. Dielectric parameters relevant to microwave dielectric heating. *Chem Soc Rev.* 1998;27:213–224.
- [10] Moriyoshi T, Ishii T, Tamai Y, Tado M. Static dielectric constants of water + ethanol and water + 2-methyl-2-propanol mixtures from 0.1 to 300 MPa at 298.15 K. *J Chem Eng Data.* 1990;35:17–20.
- [11] Hsieh C-J, Chen J-M, Li M-H. Dielectric constants of aqueous diisopropanolamine, diethanolamine, N-methyldiethanolamine, triethanolamine, and 2-amino-2-methyl-1-propanol solutions. *J Chem Eng Data.* 2007;52:619–623.
- [12] Yang LJ, Yang XQ, Huang KM, Shang H, Jia GZ. Experimental and theoretic study of the dielectric properties of ethanol + methanol mixtures. *J Solution Chem.* 2010;39:473–481.
- [13] Lone BG, Undre PB, Patil SS, Khirade PW, Mehrotra SC. Dielectric study of methanol-ethanol mixtures using TDR method. *J Mol Liq.* 2008;141:47–53.

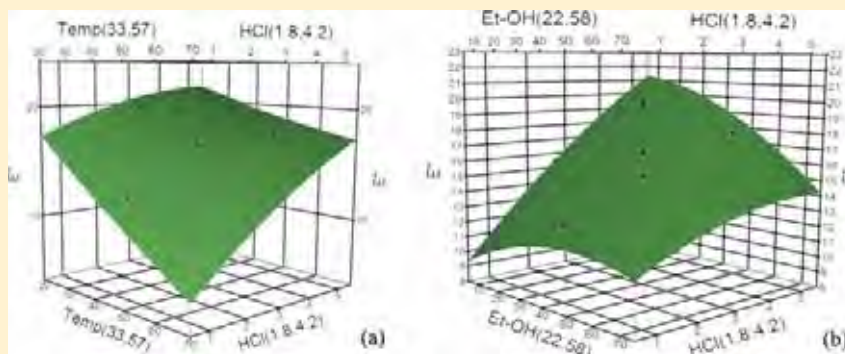


- [14] Chaudhari A, Khirade P, Singh R, Helambe SN, Narain NK, Mehrotra SC. Temperature dependent dielectric relaxation study of tetrahydrofuran in methanol and ethanol at microwave frequency using time domain technique. *J Mol Liq.* 1999;82:245–253.
- [15] Mohsen-Nia M, Amiri H, Jazi B. Dielectric constants of water, methanol, ethanol, butanol and acetone: measurement and computational study. *J Solution Chem.* 2010;39:1–8.
- [16] Datta AK, Sumnu G, Raghavan GSV. Dielectric properties of foods. Boca Raton, FL: CRC Press; 2005. p. 641–723.
- [17] Tang J. The microwave processing of foods. Dielectric Properties of Foods. Cambridge (UK): Woodhead Publishing Limited; 2005.
- [18] Liao X, Raghavan G, Yaylayan V. Dielectric properties of alcohols (C1–C5) at 2450 MHz and 915 MHz. *J Mol Liq.* 2001;94:51–60.
- [19] Venkatesh M, Raghavan G. An overview of dielectric properties measuring techniques. *Can Biosys Eng.* 2005;47:7.15–7.30.
- [20] Nemes SM. Practical methods for lignans quantification. [PhD Thesis Bioresource Eng.]. Montreal: McGill University; 2012.
- [21] Myers RH, Montgomery DC, Anderson-Cook CM. Response surface methodology: process and product optimization using designed experiments. Hoboken, NJ: John Wiley & Sons Inc; 2009.

# Dielectric Properties of Concentration-Dependent Ethanol + Acids Solutions at Different Temperatures

Winy Routray and Valerie Orsat\*

Bioresource Engineering Department, Macdonald Campus, McGill University, Ste-Anne-de-Bellevue, Quebec, Canada, H9X 3 V9



**ABSTRACT:** Dielectric properties of ethanol + HCl and ethanol + acetic acid heated at different temperature levels were recorded at microwave frequencies ranging from 0.5 GHz to 6 GHz, and the corresponding dissipation factor and depth of penetration of the solutions were calculated and analyzed using response surface methodology for the factors of temperature level and ethanol and acid (HCl and acetic acid) concentration generated using a central composite design. In case of mixtures with acetic acid, ethanol concentration and temperature were found to be the significant factors; however in the case of mixtures with HCl, HCl concentration was also found to be a significant factor contributing to all analyzed responses. The heating rate was interpreted in terms of the different factors based on their effects on dissipation factor and depth of penetration. Overall, with an increase in the concentration of ethanol and at lower temperature levels, the microwave heating rate was interpreted to be highest at 2.45 GHz. Among the acids, HCl was the only contributing factor and was interpreted to positively affect the heating rate; that is, with an increase in HCl concentration, the heating rate was observed to increase at 2.45 GHz.

## INTRODUCTION

Microwave heating has been found to be an energy-efficient technology; hence it has been increasingly applied in food processing and several analytical studies. Extraction is one of those analytical studies where microwave has been applied and has been found to be very effective with different solvents.<sup>1–3</sup> With emphasis on more energy-efficient methodologies and environmentally friendly solvents, a combination of microwave with proper solvents can significantly decrease the energy consumption, time of analytical processes, and solvent consumption. Some of the common solvents used for phytochemical extractions and other analytical methods are ethanol,<sup>4</sup> acetic acid, and hydrochloric acid (HCl),<sup>5</sup> which have also been applied in some microwave based analytical methods.<sup>6,7</sup>

Some of the factors which govern the efficiency of the microwave assisted analytical methods are the behavior of the material subjected to microwave environment, which depends on the dielectric properties of the material. Some of the important dielectric properties of solvents include dielectric constant ( $\epsilon'$ ) and dielectric loss ( $\epsilon''$ ), from which other parameters such as dissipation factor ( $\tan \delta$ ) and depth of penetration ( $D_p$ ) can be determined.<sup>8</sup> Dielectric constant and  $\epsilon''$  are related to the relative complex permittivity, which is associated with Maxwell's

equations which govern the dissemination of microwaves in the materials in the microwave environment. The dielectric constant of a material relates to the ability of a material to store electrical energy in a microwave environment. Dielectric loss and  $\tan \delta$  (which is the ratio of  $\epsilon''$  to  $\epsilon'$ ) are related to the ability of the material to store and convert electromagnetic energy into thermal energy. With increasing distance, microwaves attenuate, and the depth of penetration is the term used for the distance over which 63% of the microwave power is dissipated; hence it provides information regarding the “relative microwave absorbing characteristics of the materials”.<sup>9</sup>

The Debye model has been reported to potentially predict the dielectric properties of mixture systems (which is done in terms of relaxation time), however as discussed in a previous study, modification of the Debye equation for determination of dielectric properties has been emphasized recently.<sup>10</sup> Direct measurement of dielectric properties of mixture solutions has attracted attention recently which can be further used to explain the behavior and heating properties of mixture systems in a microwave field.<sup>10,11</sup> Hence the measurement of dielectric

Received: January 4, 2013

Accepted: May 1, 2013

Published: May 14, 2013

properties and study of parameters such as  $\tan \delta$  and  $D_p$  are reported to provide the necessary information regarding the expected overall efficiency of a process and its rate of heating.

Acids and ethanol are in general polar compounds and hence can be potential solvents for effective heating using dielectric heating (as their relaxation process occurs in the microwave frequency range).<sup>8</sup> There are several reports on the studies of dielectric properties of alcohols; however there are no reports on the dielectric properties of combinations of alcohol and acid. It was observed that, using appropriate computer aided statistical designs and analysis, microwave heating behavior of different solutions could also be predicted for different microwave frequencies which could be further used for designing microwave reactors operating at other frequencies (other than 2.45 GHz) and prediction of corresponding behavior of the substance in the microwave environment. Hence, it was concluded that, to predict the pattern of behavior of ethanol + acid combinations, appropriate further studies are required which would be useful in calculating and interpreting the performance of different analytical methods involving microwave application with novel solvents.

For the current study, the dielectric properties of solutions with different concentrations of ethanol + HCl or acetic acid (with relatively higher molarity level) were measured, at different temperature levels. Dissipation factor and  $D_p$  were calculated, and the rate of heating and microwave absorbing characteristics of the mixtures were interpreted and discussed. The pattern of variation of dielectric properties with respect to frequency was also analyzed. Finally the heating behavior of some of the combinations was observed by introducing the mixtures in the microwave reactor and was correlated with the results of dielectric measurements and analyses.

## EXPERIMENTAL SECTION

**Materials and Equipment.** Anhydrous ethyl alcohol (Commercial Alcohols, Brampton, Ontario, Canada), HCl (Fisher Scientific, Fair Lawn, New Jersey, USA), and glacial acetic acid (Fischer Scientific, Nepean, Ontario, Canada) (Table 1) were used for preparation of the respective solutions with

**Table 1. Description of Chemicals Used for the Study**

chemical used	source	purity	purification process
ethanol	Commercial Alcohols	100 <sup>a</sup>	none
HCl	Fisher Scientific	> 99.7 <sup>b</sup>	none
acetic acid	Fisher Scientific	36.5 to 38 <sup>b</sup>	none

<sup>a</sup>The purity of chemical represents mole fraction. <sup>b</sup>The purity is in mass fraction, w/w %.

water [HPLC grade water prepared using Simplicity Water Purification System (Millipore, USA)] at room temperature (293.15 to 298.15 K) of different concentrations, decided according to the central composite design (CCD) generated.

**Table 2. Independent Variables and Their Coded and Actual Values Used for the Central Composite Design<sup>a</sup>**

independent variables	unit	coded levels				
		-1.682	-1	0	1	1.682
temperature ( <i>t</i> , temp)	°C	24.8	33	45	57	65.2
ethanol concentration (EtOH, % v/v)	cm <sup>3</sup> /100 cm <sup>3</sup>	9.72	22	40	58	70.3
acid concentration (molarity (M)) (acetic acid/hydrochloric acid)/(AA/HCl)	mol·dm <sup>-3</sup>	0.98	1.8	3	4.2	5.02

<sup>a</sup>The generated CCD will be different if temperature is considered as Kelvin and acid concentration in molality.

**Statistical Design and Analysis.** The experimental design used for the study was a central composite design with uniform precision. The three factors considered for this study included temperature level (°C), ethanol concentration (% v/v, cm<sup>3</sup>/100 cm<sup>3</sup>), and acid concentration (mol·dm<sup>-3</sup>) (Table 2). The CCD generated for the experiment would be different if temperature was considered in K and acid concentration in molality (mol·kg<sup>-1</sup>), leading to variations in the observation and analysis, and corresponding interpretation. The corresponding values of temperature in K and acid concentrations in molality have been reported in Table 3. Two types of acids were considered, namely, HCl and acetic acid. Separate CCDs were developed for each acid. Hence three factors varying at five different levels (-1.682, -1, 0, 1, 1.682) were used to devise a CCD experimental design (Table 2). Twenty experimental runs (the total number of experimental combinations depends on the number of factors) were generated, including eight factorial, six axial, and six central points (Table 3). Temperature levels were varied between 24.8 °C (297.95 K) and 65.2 °C or 338.35 K (with a variation maintained between ± 0.5 °C or K), alcohol concentration varied from 9.72 % v/v and 70.3 % v/v, and acid concentration ranged between 0.98 mol·dm<sup>-3</sup> to 5.02 mol·dm<sup>-3</sup> (0.97 mol·kg<sup>-1</sup> to 4.65 mol·kg<sup>-1</sup> in the case of HCl and 0.98 mol·kg<sup>-1</sup> to 4.95 mol·kg<sup>-1</sup> in the case of acetic acid). JMP software version 8 (SAS Institute Inc., Cary, NC, USA) was used for building the experimental design and for the analysis of the collected data. The experimental combinations, other than the six central points, were replicated thrice. Hence, in total 15 experimental combinations and 48 experiments were conducted for each acid.

Graphs of the  $\epsilon'$ ,  $\epsilon''$ , and  $\tan \delta$  were generated using Microsoft excel for all the 15 experimental combinations taking the average value of the observations, and the variation of these properties and parameters with respect to different microwave frequencies was studied and has been discussed in the following sections.

In the later part of the study, response surface methodology (RSM) was used to analyze the data for CCD with 20 experimental runs. Average values of the responses for  $\epsilon'$ ,  $\epsilon''$ ,  $\tan \delta$ , and  $D_p$  were calculated for the eight factorial points and six axial points (as all of these points were replicated thrice). Values of responses for  $\epsilon'$ ,  $\epsilon''$ ,  $\tan \delta$ , and  $D_p$  at the six central points were combined with the average values of responses at other experimental points mentioned above to analyze the result using RSM methodology. Individual (linear), quadratic, and combined or interactive (bilinear) effects were interpreted to determine which factors are the most significant and the pattern of their effect. Regression models were developed for all of the responses and validity of these models were analyzed using ANOVA. The response function for  $\epsilon'$ ,  $\epsilon''$ ,  $\tan \delta$ , or depth of penetration (*Y*) was partitioned into linear, quadratic, and bilinear components:

**Table 3. Central Composite Showing the Different Combination of the Temperature Levels, Different Alcohols, and Different Types of Acids with Different Concentrations<sup>a</sup>**

treatment no.	temperature		ethanol concentration (% v/v, cm <sup>3</sup> /100 cm <sup>3</sup> )	acid concentration			
	t/°C	T/K		molarity/mol·dm <sup>-3</sup>			
			HCl	acetic acid			
1	33	306.15	22	1.8	1.75	1.79	factorial points
2	33	306.15	22	4.2	3.94	4.15	
3	33	306.15	58	1.8	1.75	1.79	
4	33	306.15	58	4.2	3.94	4.15	
5	57	330.15	22	1.8	1.75	1.79	
6	57	330.15	22	4.2	3.94	4.15	
7	57	330.15	58	1.8	1.75	1.79	
8	57	330.15	58	4.2	3.94	4.15	
9	24.8	297.95	40	3	2.87	2.97	axial points
10	65.2	338.35	40	3	2.87	2.97	
11	45	318.15	9.72	3	2.87	2.97	
12	45	318.15	70.3	3	2.87	2.97	
13	45	318.15	40	0.98	0.97	0.98	
14	45	318.15	40	5.02	4.65	4.95	
15–20	45	318.15	40	3	2.87	2.97	center points

<sup>a</sup>T = temperature; EtOH = ethanol concentration. The generated CCD will be different if temperature is considered as Kelvin and acid concentration in molality.

$$Y = \beta_0 + \sum_{i=1}^k B_i X_i + \sum_{i=1}^k B_{ii} X_i^2 + \sum_{i>j} B_{ij} X_i X_j \quad (1)$$

where  $\beta_0$  represents the model intercept;  $B_i$ ,  $B_{ii}$ , and  $B_{ij}$  denote the regression coefficients of the linear, quadratic, and bilinear effects, respectively;  $k$  is the number of independent variables or factors, which is 3 in this case; and  $X_i$  and  $X_j$  are the coded independent variables. Table of parameter estimates of different significant factors with their regression coefficients (parameter estimates) were prepared using analysis of variance for the CCD-RSM analysis, where significance levels were evaluated based on the  $F$  value at the probability ( $p$ ) of 0.001, 0.01, or 0.05. Response surfaces, which are the three-dimensional figures were generated based on the response functions.  $X$  and  $Y$  axes represented any two significant factors, and  $Z$  axis represented any one response. The response function (regression model) developed and response surface fitted over "the region around the current selected conditions" can be used in predicting the response resulting from any readjustment of the independent factors (temperature level, ethanol concentration, acid concentration).<sup>12</sup>

**Preparation of Samples and Dielectric Properties Measurement.** A total volume of 20 mL of mixture solution was prepared in glass tubes consisting of 19.8 mL of different concentrations of ethanol and 0.2 mL of acid solutions with different molarities, which was heated up to the required temperature determined by the CCD design, and then the dielectric properties were measured. The solutions were heated using an Isotemp heater (2001FS, Thermofisher Scientific, USA). Then the open-ended high-temperature (flat) probe (85070D, Agilent Technologies, USA) was inserted into the solution which, in conjunction with the ECal module (N4693A, Agilent Technologies, USA), was attached to the dielectric measurement equipment (Agilent S Parameter Network Analyzer, 8722ES, Agilent Technologies, USA) which was used for the measurement of the dielectric properties which included  $\epsilon'$  and  $\epsilon''$ . As a part of calibration of the probe, the dielectric properties of HPLC grade water at 298.15 K were also measured.

The collected data was used for the calculation of the  $\tan \delta$  and  $D_p$ . The dissipation factor was calculated using the following formula:

$$\tan \delta = \frac{\epsilon''}{\epsilon'} \quad (2)$$

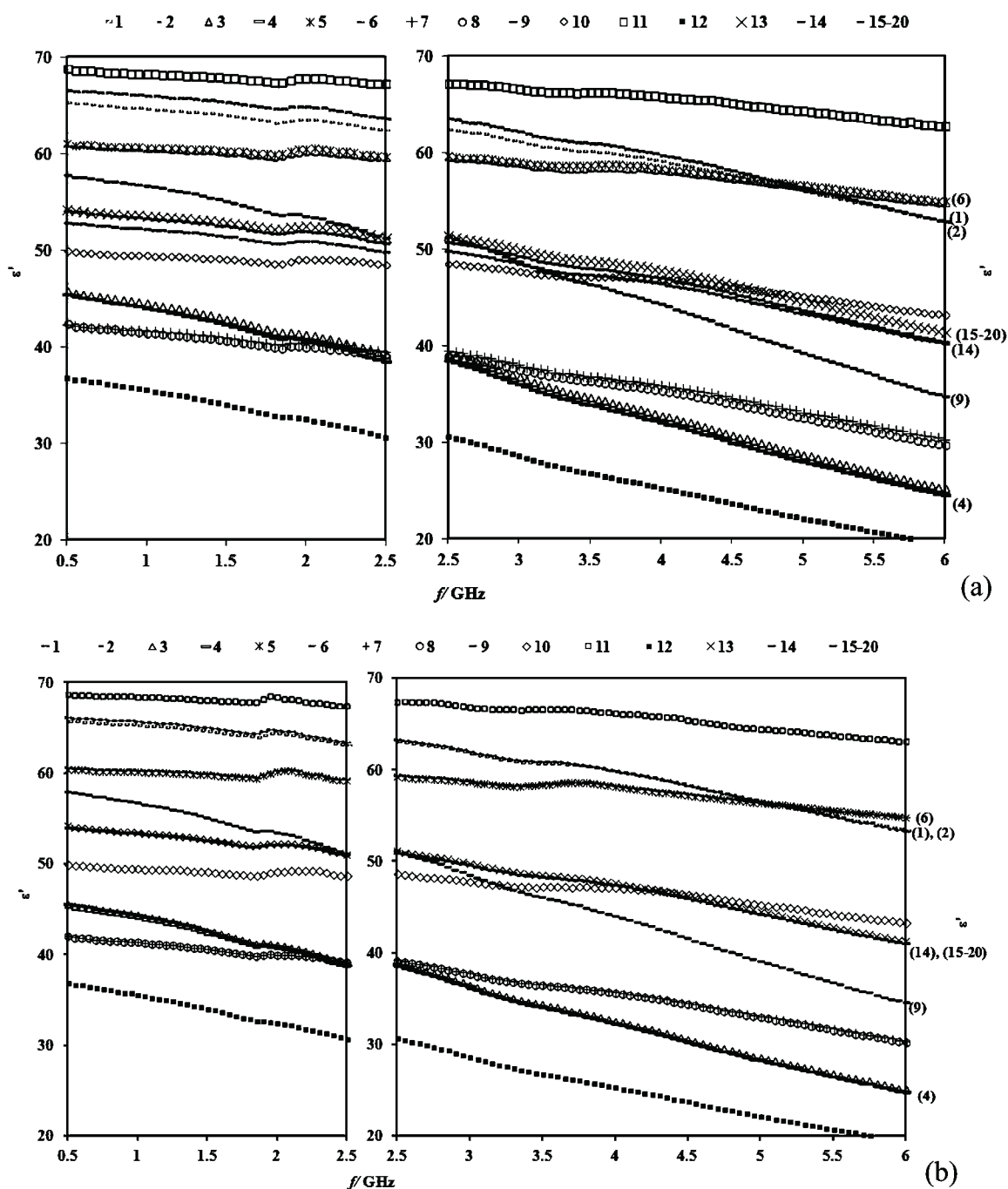
where  $\tan \delta$  is the dissipation factor,  $\epsilon''$  is the dielectric loss, and  $\epsilon'$  is the dielectric constant.

The depth of penetration is calculated as:

$$D_p = \frac{\lambda_0}{2\pi\sqrt{2\epsilon'}} (\sqrt{1 + (\epsilon''/\epsilon')^2} - 1)^{1/2} \quad (3)$$

where  $D_p$  is the penetration depth,  $\lambda_0$  is the wavelength of microwaves in free space (0.122 m at 2450 MHz microwave frequency), and all other symbols represent the parameters mentioned for eq 2.

**Measurement of Temperature Variation with Combinations of Ethanol and Acids in the Microwave Reactor.** During this study the rate of temperature increase, in the microwave reactor, of ethanol combined with higher acid concentrations was observed using a fiber-optic probe to observe the effect of higher concentration of acids along with other factors on the microwave heating behavior of the alcohol and acid combinations. A monomode microwave system (Star System 2, CEM, Mathews, USA) with a nominal power of 800 W at 2450 MHz frequency was used for heating the mixture solutions of a 40 mL volume (0.4 mL of acid solution + 39.6 mL of ethanol solution) which were taken in a round bottomed tubular vessel made of borosilicate glass. The actual power of the system has been reported as 710.5 W via standard calibration,<sup>13</sup> and the solution was heated at 10 % actual power level for 240 s. At each 30 s interval, the temperatures of solutions prepared by mixing 22 cm<sup>3</sup>/100 cm<sup>3</sup> or 58 cm<sup>3</sup>/100 cm<sup>3</sup> ethanol with 1.8 mol·dm<sup>-3</sup> or 4.2 mol·dm<sup>-3</sup> acid (HCl or acetic acid) were measured and graphical plots with temperature versus time were generated to observe the rate of increase of temperature in each case. The highest temperature achieved was less than 90 °C or 363.15 K,



**Figure 1.** Variation of  $\epsilon'$  with different combinations of (a) HCl acid and ethanol concentrations and (b) acetic acid and ethanol concentrations at different temperatures with respect to frequency ( $f/\text{GHz}$ ). Some of the plots are overlapped; hence the plots which are versions of dotted lines or dashed lines are mentioned in the side in terms of their experimental combination numbers. The plots which are very closely plotted and cannot be visualized separately have been mentioned individually and separated using commas as the experimental combination number at the end of the plot. When two plots are coinciding at one end point, the top number is for the upper plot, and the bottom number is for the lower plot.

which prevented solvent loss and any accident due to high temperature of solvent in the microwave system.

## RESULTS AND DISCUSSION

**Graphical Analysis of the Variation of Dielectric Properties. Dielectric Constant.** As observed from Figure 1, the  $\epsilon'$  decreased with the increase in frequency for all of the combinations for both the acids. However, the rate of change in

the  $\epsilon'$  was different for different combinations. For the frequency range chosen for this study, the  $\epsilon'$  is continuously decreasing, which generally happens because of the phase lag between the dipole alignment and the electric field. It also implied that this frequency range is above the relaxation frequency of the combinations/mixture solutions used for this study. It was observed that there is a difference in the pattern of variation of  $\epsilon'$  with respect to different combinations of ethanol + acid



Table 4. Average Value of  $\epsilon'$ ,  $\epsilon''$ ,  $\tan \delta$ , and  $D_p$  at Various CCD Combinations of Different Ethanol Concentrations, Temperature Levels, and HCl and Acetic Acid Concentrations<sup>a</sup>

expt. comb. no.	HCl				acetic acid			
	$\epsilon'$	$\epsilon''$	$\tan \delta$	$D_p/m$	$\epsilon'$	$\epsilon''$	$\tan \delta$	$D_p/m$
1	63.65 ± 0.14	16.04 ± 0.48	0.2520 ± 0.0076	0.0097 ± 0.0003	63.50 ± 0.45	12.36 ± 0.20	0.1947 ± 0.0035	0.0126 ± 0.0002
2	62.63 ± 0.36	19.71 ± 0.73	0.3147 ± 0.0118	0.0079 ± 0.0003	63.23 ± 0.44	12.13 ± 0.09	0.1918 ± 0.0020	0.0128 ± 0.0001
3	39.46 ± 0.26	15.98 ± 0.41	0.4050 ± 0.0107	0.0078 ± 0.0002	39.23 ± 0.79	14.46 ± 0.30	0.3687 ± 0.0107	0.0086 ± 0.0002
4	38.77 ± 0.25	18.04 ± 0.37	0.4654 ± 0.0100	0.0069 ± 0.0001	39.04 ± 0.70	14.66 ± 0.18	0.3755 ± 0.0082	0.0084 ± 0.0001
5	59.83 ± 0.22	11.99 ± 0.43	0.2004 ± 0.0072	0.0126 ± 0.0005	59.30 ± 0.55	7.56 ± 0.26	0.1274 ± 0.0045	0.0198 ± 0.0007
6	59.44 ± 0.27	17.80 ± 1.08	0.2996 ± 0.0182	0.0085 ± 0.0005	59.43 ± 0.31	7.69 ± 0.23	0.1294 ± 0.0039	0.0195 ± 0.0006
7	39.67 ± 0.03	11.79 ± 0.63	0.2971 ± 0.0159	0.0105 ± 0.0006	39.34 ± 0.54	9.52 ± 0.63	0.2421 ± 0.0162	0.0129 ± 0.0009
8	39.12 ± 0.05	14.74 ± 0.66	0.3768 ± 0.0170	0.0084 ± 0.0004	39.29 ± 0.65	9.59 ± 0.29	0.2440 ± 0.0084	0.0128 ± 0.0004
9	51.57 ± 0.17	19.83 ± 0.29	0.3845 ± 0.0058	0.0072 ± 0.0001	51.40 ± 0.16	17.21 ± 0.27	0.3348 ± 0.0054	0.0082 ± 0.0001
10	48.64 ± 0.35	13.16 ± 0.69	0.2705 ± 0.0144	0.0104 ± 0.0005	48.74 ± 0.30	7.47 ± 0.20	0.1532 ± 0.0042	0.0182 ± 0.0005
11	67.28 ± 0.17	16.51 ± 0.72	0.2454 ± 0.0107	0.0097 ± 0.0004	67.48 ± 0.24	7.88 ± 0.30	0.1168 ± 0.0045	0.0203 ± 0.0008
12	30.87 ± 0.46	14.05 ± 0.04	0.4551 ± 0.0069	0.0079 ± 0.0001	30.88 ± 0.24	11.92 ± 0.25	0.3861 ± 0.0086	0.0092 ± 0.0002
13	51.49 ± 0.40	12.75 ± 0.36	0.2476 ± 0.0073	0.0110 ± 0.0003	51.24 ± 0.22	11.34 ± 0.14	0.2214 ± 0.0030	0.0123 ± 0.0002
14	50.03 ± 0.67	18.69 ± 1.09	0.3738 ± 0.0224	0.0075 ± 0.0004	51.11 ± 0.02	11.42 ± 0.51	0.2234 ± 0.0099	0.0123 ± 0.0005
15–20	50.91 ± 0.41	16.27 ± 0.46	0.3197 ± 0.0093	0.0086 ± 0.0002	51.12 ± 0.31	11.35 ± 0.61	0.2220 ± 0.0121	0.0124 ± 0.0007

<sup>a</sup> $\epsilon'$  = dielectric constant,  $\epsilon''$  = dielectric loss,  $\tan \delta$  = dissipation factor,  $D_p$  = depth of penetration (m). For all of the cases reported above, the reported uncertainty is based on the standard uncertainty multiplied by a coverage factor  $k = 2$ , providing a level of confidence of approximately 95 %.

concentrations along with temperature levels. However the pattern of variation of  $\epsilon'$  for HCl is very similar to acetic acid at the same ethanol, acid, and temperature levels for both acids, and values of  $\epsilon'$  range between 20 to 70 for both the acids within the microwave frequency range of 0.5 GHz and 6 GHz (which is less than dielectric constant of water (above 70) for this frequency range at 298.15 K). Hence, it can be interpreted that temperature level and ethanol concentration are the most significant factors affecting  $\epsilon'$  for combinations of ethanol with both strong and weak acids. For acetic acid the plots for combinations 1 and 2; 3 and 4; 5 and 6; 7 and 8; 13 to 20 have been found to be almost overlapping; however in the case of HCl, the plots are slightly apart from each other (not entirely overlapping each other). The average values of different responses for various combinations have been summarized in Table 4, which also supports the observation regarding the values of  $\epsilon'$  at constant temperature and ethanol concentration.

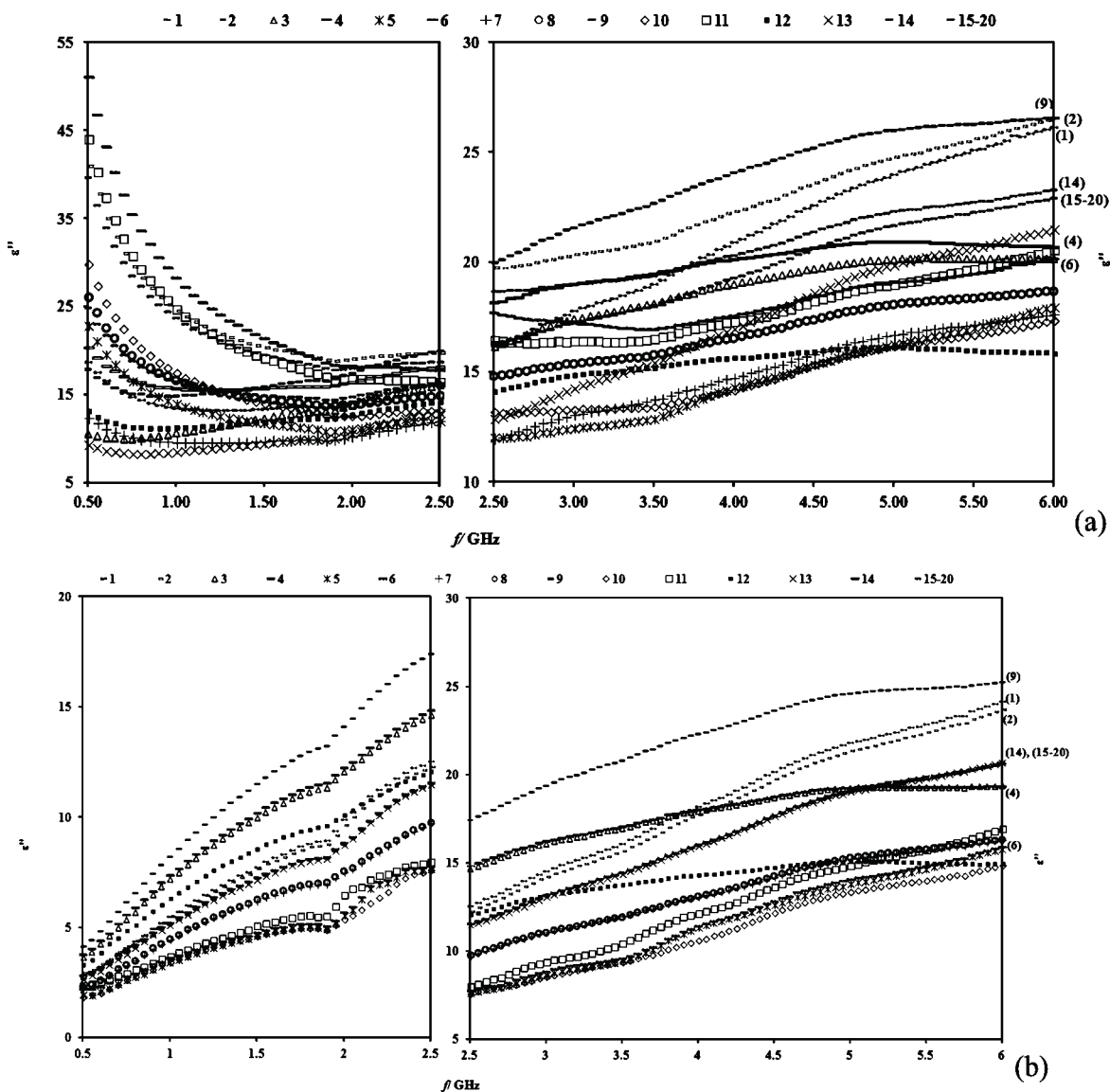
The plot for the 11th combination is highest as seen in Figure 1 for both the cases. This combination has lowest ethanol concentration. Also the 12th combination, with the highest concentration of ethanol, is lowest in Figure 1 for both the acids. This observation implies that, to obtain highest value of  $\epsilon'$ , the lowest concentration of ethanol is suitable at all of the microwave frequencies. The plot for the 10th combination and 9th combination are intersecting for both acids at a frequency between 3 GHz and 3.5 GHz; also the rate of decrease of the  $\epsilon'$  for the 10th combination (which is at higher temperature) is lower than for the 9th combination. Hence it implies that the effect of temperature is also affected by the frequency. In cases, where new frequency levels are tried, dielectric property measurement can be very helpful to understand the behavior (whether it increases or decreases and for what reasons) of the mixture solutions.

In all other cases with combinations with same concentration of ethanol and acid and different temperature levels, a similar observation can be made. The frequency at which the intersection of the plots is taking place is decreasing with the increase in the concentration of ethanol. However at 2.45 GHz, it

can be observed that with increase in temperature for combinations with similar concentrations, the  $\epsilon'$  is decreasing. Also the values of the  $\epsilon'$  for both the acids lie in the same range; hence  $\epsilon''$  can be used to interpret the heating rate. Overall, from Figure 1 and Table 4, it can be concluded that, with increase in ethanol concentration and temperature levels, the  $\epsilon'$  decreases at 2.45 GHz, which will be further analyzed in the following sections using RSM.

**Dielectric Loss.** In Figure 2, it can be observed that the graphical plots for  $\epsilon''$  varied with the acid type. For acetic acid, the  $\epsilon''$  increases with the increase in frequency from 0.5 GHz to 6 GHz; however for HCl it first decreases between the frequency range of 0.5 GHz and 1.5 GHz or 2 GHz (varies with different combinations), and then increases with the increase in frequency. However the pattern of variation varies with different ethanol and acid concentrations. The dielectric loss value varies between 5 and 55 for HCl, and for acetic acid it varies from values lower than 5 to values higher than 25 (whereas for HPLC grade water at 298.15 K, it varies from 1.5 to 21.5 for this range of frequency, which is nearer to acetic acid combinations and much less than HCl combinations). The value of  $\epsilon''$  for combinations with acetic acid is very low at lower frequencies, which might make their heating rate very low and not useful. At lower frequencies, the values of  $\epsilon''$  for mixtures with HCl are higher than the values of  $\epsilon''$  at higher microwave frequencies, and the effect of various factors is more prominent at lower frequencies because of the range of the values. However, the  $\epsilon''$  values for the combinations with acetic acids are much lower than that of HCl at the same frequency levels, and also, at the lowest frequency, the plot for different combinations are comparatively in a very small range, which can be interpreted as a less significant effect of factors at these lower frequency levels for combinations with acetic acid. From the graphical plot it is also evident that dielectric value variation and values at different frequencies are visibly affected by the type of acid.

For acetic acid, at both lower and higher frequencies the 9th combination (with lowest temperature) had the highest  $\epsilon''$  value. In case of HCl, the ninth combination obtained highest values at

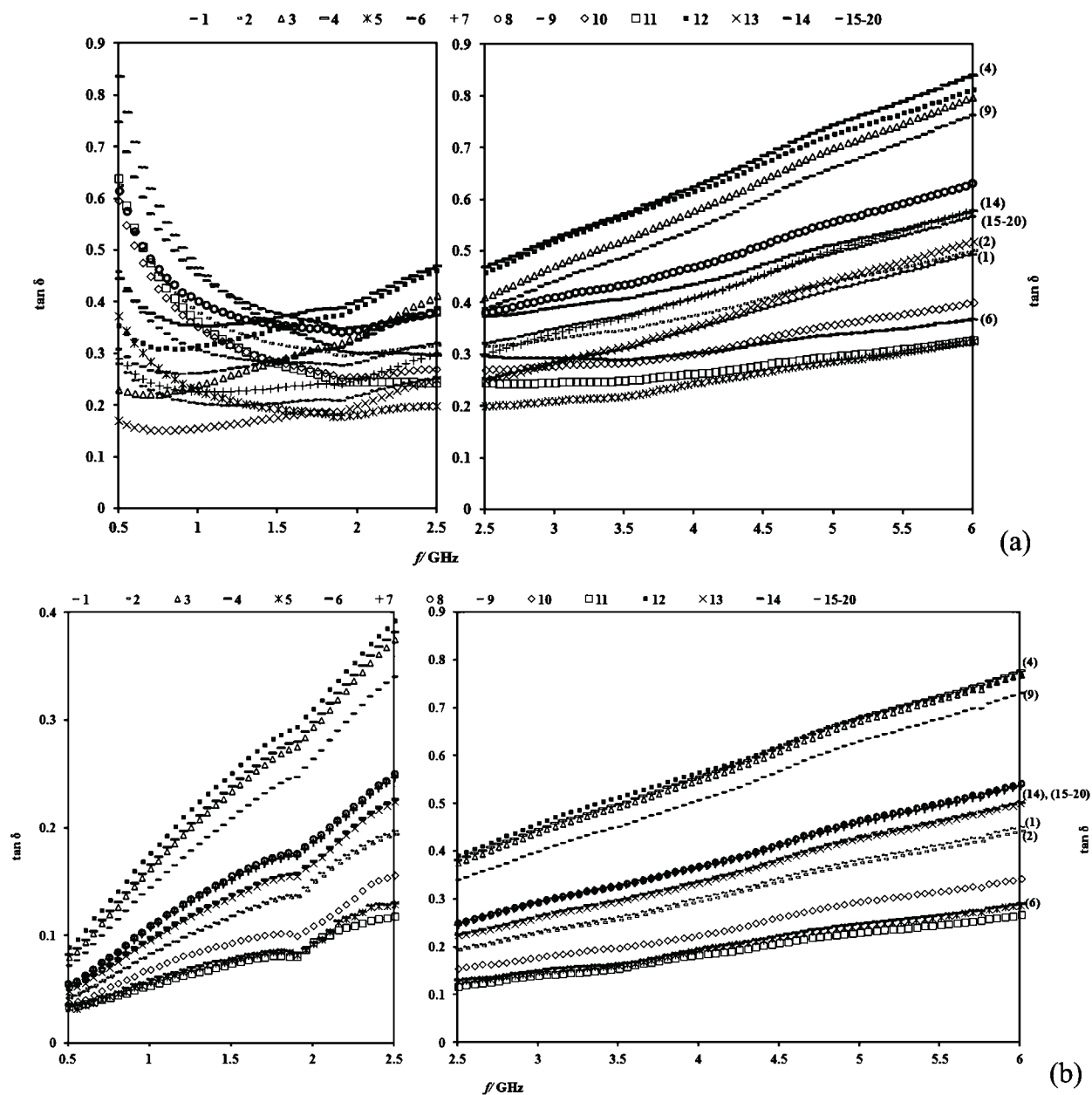


**Figure 2.** Variation of  $\epsilon''$  with respect to different combinations of (a) HCl acid and ethanol and (b) acetic acid and ethanol at different temperatures with respect to frequency ( $f/\text{GHz}$ ). Some of the plots are overlapped; hence the plots which are versions of dotted lines or dashed lines are mentioned in the side in terms of their experimental combination numbers. The plots which are very closely plotted and cannot be visualized separately have been mentioned individually and separated using commas as the experimental combination number at the end of the plot. When two plots are coinciding at one end point, the top number is for the upper plot, and the bottom number is for the lower plot.

all frequencies except some very few (some frequencies around 2 GHz). In case of combinations with acetic acid, the 10th combination (with highest temperature among all the combinations) is a combination with one of the lowest values of  $\epsilon''$  at all frequency levels. However for HCl, the 10th combination has higher  $\epsilon''$  values than many other combinations at very low frequencies (below 1.5 GHz). Also it (10th combination with HCl) is not having the lowest values of  $\epsilon''$  at higher frequencies as well. The plot for 10th combination intersects plot for combinations 5th, 7th, and 12th, which can be interpreted as higher temperature effect and effect of interaction of all the factors in case of HCl, which will be verified and discussed in the next sections with RSM.

For acetic acid, the various combinations with same ethanol concentration and temperature levels, such as 1 and 2; 3 and 4; 5 and 6; 7 and 8; 13 to 20, the plots are highly overlapping (Figure

2b); however in the case of HCl, the plots of these combinations are far apart. This implies that the effect of variation of concentration for acetic acid is negligible; however, for HCl, there is a significant effect of variation of concentration. From Figure 2a it can be observed that with an increase in the HCl concentration the  $\epsilon''$  increases and hence may increase the heating rate. This might be due to the significant increase in polarity as HCl is a strong acid and at higher concentrations it has a significant effect. For HCl the values of  $\epsilon''$  seem to decrease with an increase in ethanol concentration and temperature level at 2.45 GHz (Table 4); for acetic acid, as interpreted from Table 4 and Figure 2b, it seems that  $\epsilon''$  decreases with an increase in temperature and a decrease in ethanol concentration at 2.45 GHz; however the intersection of the different plots for both the acids demonstrates the interaction of ethanol concentration with other factors, which will be analyzed and confirmed in next



**Figure 3.** Variation of  $\tan \delta$  with different combinations of (a) hydrochloric acid and (b) acetic acid and ethanol at different temperatures with respect to frequency ( $f$ /GHz). Some of the plots are overlapped; hence the plots which are versions of dotted lines or dashed lines are mentioned in the side in terms of their experimental combination numbers. The plots which are very closely plotted and cannot be visualized separately have been mentioned individually and separated using commas as the experimental combination number at the end of the plot. When two plots are coinciding at one end point; the top number is for the upper plot, and the bottom number is for the lower plot.

sections. The variation of heating rate with variation of temperature and ethanol concentrations cannot be explained well using this graphical analysis, as there are many interaction effects; hence the interpretation of heating rate in terms of  $\epsilon''$  at 2.45 GHz for this study will be done following RSM, which emphasizes the importance of this methodology as well.

**Dissipation Factor.** Figure 3 presents the variation of  $\tan \delta$  for various combinations with different ethanol and acid concentrations. The  $\tan \delta$ , in case of acetic acid, increases with an increase in frequency for all the cases; however the rates of increase vary. The heating rate for the same combination can be expected to increase with an increase in frequency. The  $\tan \delta$  of

combinations with HCl at higher molarity first decreases and then again increases with an increase in the frequency.

In case of acetic acid, the plots of the mixture solutions with higher concentration of ethanol, which are the 12th, 3rd, and 4th combinations, are higher than others (which includes 1st, 2nd, 14 to 20th combinations). With an increase in temperature the  $\tan \delta$  decreases (the plots for 3rd and 4th; 1st and 2nd; 9th are higher than 7th and 8th; 5th and 6th; 10th combinations, respectively), however the overlapping (or closed placement) of 3rd, 4th, with 12th, and of 5th, 6th, with 11th demonstrates a possible significant interaction effect among different factors. Yet the similarity of 3rd to 4th, 1st to 2nd, 5th to 6th, and 7th to 8th

**Table 5. Corresponding Parameter Estimates at Coded Factor Levels for the Predictive Model Derived from ANOVA for the CCD-RSM Analyses of  $\epsilon'$ ,  $\epsilon''$ ,  $\tan \delta$ , and  $D_p$  with Respect to Different Ethanol Concentrations, Temperature Levels, and HCl or Acetic Acid Concentrations<sup>a</sup>**

predictive model	parameter estimate for the source of variation in case of CCD of temperature, ethanol, and acetic acid									
	intercept	temp	EtOH	acid (AA)	temp-acid	EtOH-acid	temp-EtOH	temp-temp	EtOH-EtOH	acid-acid
$\epsilon'$ ( $R^2 = 0.9993$ , RMSE = 0.3309)*	51.1091	-0.8862*	-10.9913*				1.0466*	-0.3255 <sup>#</sup>	-0.6404*	
$\epsilon''$ ( $R^2 = 0.9681$ , RMSE = 0.6191)*	11.3565	-2.6087*	1.1193*						-0.5549 <sup>#</sup>	
$\tan \delta$ ( $R^2 = 0.9886$ , RMSE = 0.0117)*	0.2222	-0.0507*	0.0761*				-0.0161 <sup>#</sup>		0.0091 <sup>#</sup>	
$D_p$ ( $R^2 = 0.9836$ , RMSE = 0.0006)*	0.0124	0.0029*	-0.0030*				-0.0007 <sup>#</sup>		0.0008 <sup>§</sup>	
predictive model	parameter estimate for the source of variation in case of CCD of temperature, ethanol, and HCl									
	intercept	temp	EtOH	acid (HCl)	temp-acid	EtOH-acid	temp-EtOH	temp-temp	EtOH-EtOH	acid-acid
$\epsilon'$ ( $R^2 = 0.9988$ , RMSE = 0.4528)*	50.8949	-0.8334*	-10.9645*	-0.3748 <sup>#</sup>			0.9479 <sup>§</sup>		-0.5629 <sup>§</sup>	
$\epsilon''$ ( $R^2 = 0.9795$ , RMSE = 0.4595)*	16.2753	-1.8063*	-0.6682 <sup>§</sup>	1.7931*	0.3803 <sup>#</sup>	-0.5588 <sup>#</sup>			-0.3598 <sup>#</sup>	
$\tan \delta$ ( $R^2 = 0.9916$ , RMSE = 0.0087)*	0.3198	-0.0333*	0.0608*	0.0376*	0.0070 <sup>#</sup>		-0.0162 <sup>§</sup>		0.0100 <sup>#</sup>	
$D_p$ ( $R^2 = 0.9833$ , RMSE = 0.0003)*	0.0086	0.0010*	-0.0006*	-0.0011*	-0.0004 <sup>§</sup>	0.0004 <sup>#</sup>				0.0002 <sup>#</sup>

<sup>a</sup>EtOH = ethanol concentration, temp = temperature levels, acid = acid concentration level, RMSE = root-mean-square error. All of the models statistically generated had insignificant lack of fit. The CCD was created using temperature in °C and acid concentration in molarity (mol·dm<sup>-3</sup>); however the analyses and predictive models will change if temperature is considered in K and acid concentration in molality (mol·kg<sup>-1</sup>). \* represents significance at  $p < 0.0001$ , § represents significance at  $p < 0.001$ , and # represents significance at  $p < 0.05$ .

proves a negligible effect of acid concentration on the  $\tan \delta$  in the case of acetic acid.

In case of HCl, there is a possible significant effect of HCl concentration as the plot for 4th is higher than 3rd at all frequencies and 2nd is higher than 1st (Figure 3a). The different combinations are having different reactions with increasing frequency. Due to the intersection of different plots at several points and also lack of fixed pattern of variation of the plots at lower frequencies, it is difficult to interpret the effect of different factors, however it shows high interaction effects. The combinations with higher ethanol concentrations (3rd, 4th, and 12th combinations) have higher dissipation factors at higher frequencies than others. At higher frequencies the mixture solutions subjected to lower temperature levels (3rd, 4th, and 9th combinations) were having higher  $\tan \delta$  than the same mixture solutions subjected to higher temperature levels (7th, 8th and 10th, 15th combinations). Hence, in case of combinations with HCl with higher molarity, all of the factors and their interaction effects possibly have a significant effect which will be analyzed again using RSM.

**Response Surface Analysis of Different Dielectric Properties.** Dielectric constant,  $\epsilon''$ ,  $\tan \delta$ , and  $D_p$  (as calculated using eq 3) obtained for the 20 combinations (Tables 3 and 4) were analyzed using RSM. Table 5 summarizes the corresponding parameter estimates at coded factor levels for the predictive model derived from ANOVA for the CCD-RSM analyses of  $\epsilon'$ ,  $\epsilon''$ ,  $\tan \delta$ , and  $D_p$  with respect to different ethanol concentrations, temperature levels, and HCl or acetic acid concentrations. The corresponding  $R^2$  and root-mean-square error (RMSE) values for the different predictive model for  $\epsilon'$ ,  $\epsilon''$ ,  $\tan \delta$ , and  $D_p$  of ethanol + acid combinations have also been included along with the level of significance of model and different parameter estimates. The parameter estimates mentioned in Table 5 can be used for generating a predictive model in terms of coded factor levels which have high  $R^2$  and insignificant lack of fit (only the models satisfying the criteria have been summarized in the table).

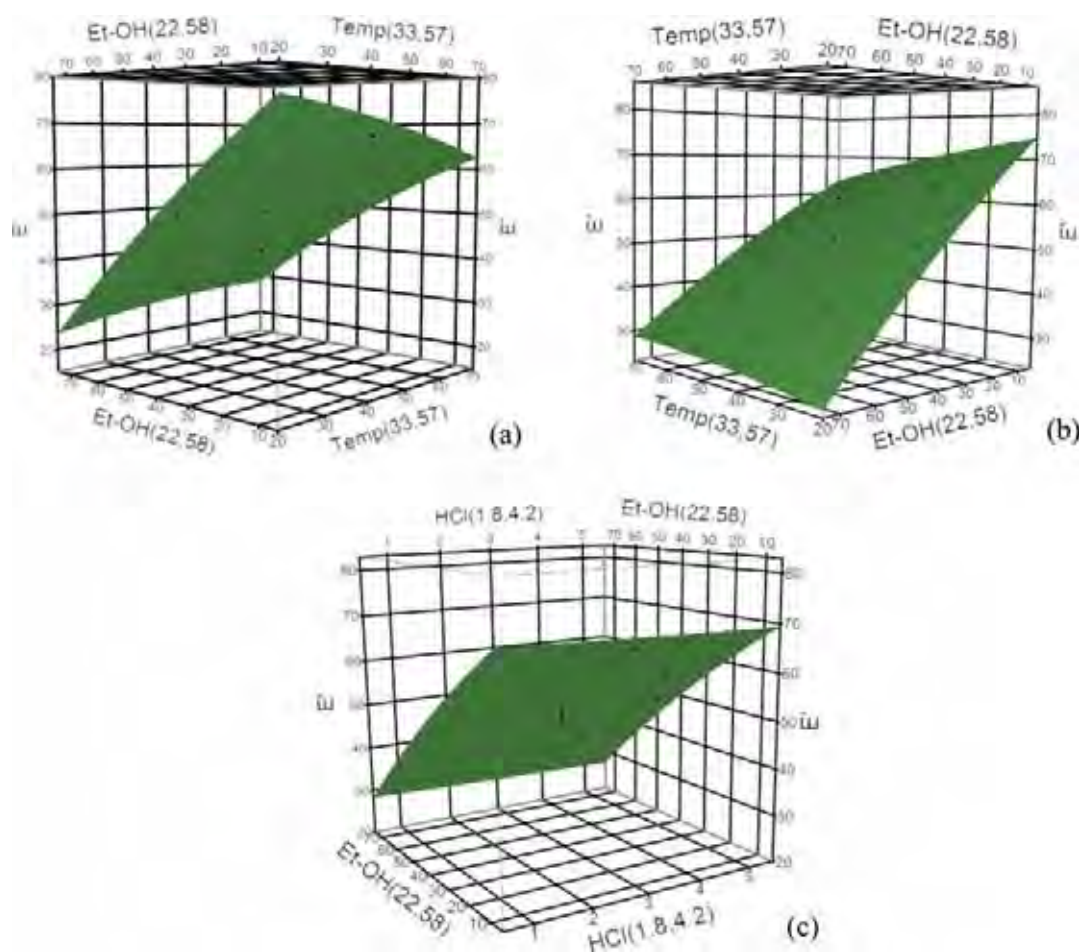
The generation of a predictive model using Table 5 has been demonstrated in case of  $\epsilon'$  of ethanol + acetic acid and can be repeated in the same way for other cases. The table of parameter estimates obtained for different factors and response surfaces generated were used to analyze and discuss the importance of various factors and their contribution in the final response.

**Dielectric Constant.** For acetic acid and ethanol combinations the influential factors included temperature and ethanol. As interpreted in section on  $\epsilon'$  in graphical analyses, acetic acid concentration does not have a significant effect. A predictive model with linear (temperature and ethanol concentration), bilinear (temperature-ethanol concentration), and quadratic effect (temperature<sup>2</sup> and ethanol<sup>2</sup>) was generated using the parameter estimates summarized in Table 5, which was found significant at  $\alpha < 0.0001$  (as indicated by sign "\*" in Table 5) with a  $R^2$  value of 0.9993 and insignificant lack of fit indicating the accuracy and adequacy of the predictive model representing the system. Predictive model of  $\epsilon'$  of acetic acid + ethanol combinations can be presented in terms of significant factors as

$$\begin{aligned} \epsilon'_{AA} = & 51.1091 - 0.8862 \cdot \text{temp} - 10.9913 \cdot \text{EtOH} \\ & + 1.0466 \times \text{temp} \cdot \text{EtOH} - 0.3255 \cdot \text{temp}^2 \\ & - 0.6404 \cdot \text{EtOH}^2 \end{aligned} \quad (4)$$

In terms of the value of parameter estimates for the linear effects, it can be observed that ethanol was a more significant factor than temperature level, and with the increase in both factors, the  $\epsilon'$  of acetic acid + ethanol mixtures decreased (as interpreted in the previous section on  $\epsilon'$ ). From Figure 4a it can be observed that, with increase in ethanol concentration and temperature levels, the  $\epsilon'$  decreased; however in the three-dimensional graph, where both ethanol concentration and temperature are high, the  $\epsilon'$  has a comparatively higher value than at the point with lowest temperature and highest ethanol concentration. This can be explained by the positive value of the parameter estimate for the bilinear term in the predictive model,





**Figure 4.** Response surface plot representing effect of (a) ethanol concentration ( $\text{cm}^3/100 \text{ cm}^3$ , EtOH) and temperature levels ( $^{\circ}\text{C}$ , temp) on dielectric constant of different ethanol and acetic acid concentration combinations at various temperature levels, (b) ethanol concentration ( $\text{cm}^3/100 \text{ cm}^3$ , EtOH) and temperature levels ( $^{\circ}\text{C}$ , temp) on dielectric constant of various HCl and ethanol concentration combinations at various temperature levels, and (c) HCl concentration ( $\text{mol}\cdot\text{dm}^{-3}$ , HCl) and ethanol concentration ( $\text{cm}^3/100 \text{ cm}^3$ , EtOH) on dielectric constant of various HCl and ethanol concentration combinations at various temperature levels.

which implies that the interaction of temperature and ethanol concentration has a positive effect on  $\epsilon'$ ; hence at the combination with higher temperature and higher ethanol concentration, the  $\epsilon'$  was found to be higher than what is expected in terms of individual temperature and ethanol concentration increase. The quadratic terms of  $\text{temp}^2$  and  $\text{EtOH}^2$  are used for interpreting if there is a point of maximum or minimum for the response in the range of factors considered for the study. In this case, according to Table 5, both  $\text{temp}^2$  and  $\text{EtOH}^2$  have negative parameter estimate value which implies a point of maximum of response for both these factors (Figure 4a). These interpretations agree with the discussion in previous section on  $\epsilon'$  in graphical analysis.

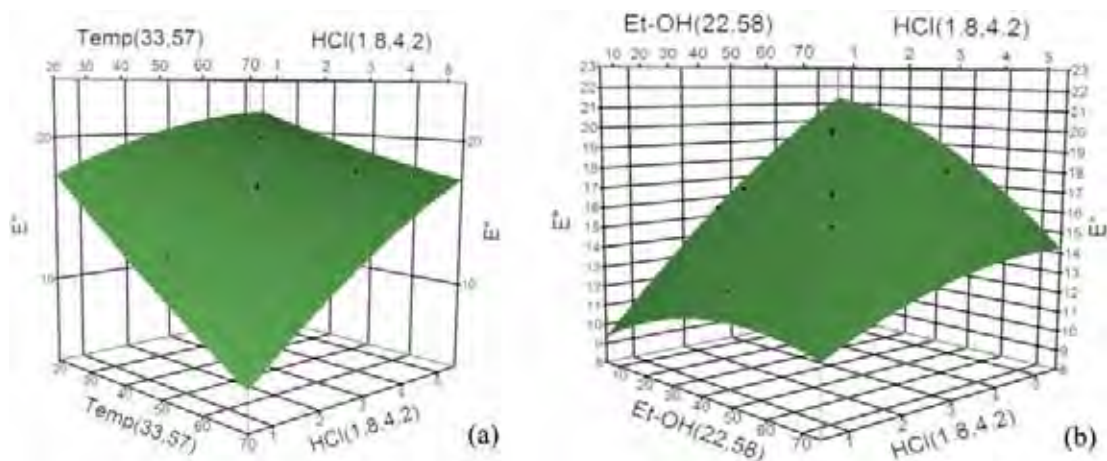
In the case of a combination with HCl + ethanol, the significant factors included linear effects of temperature, ethanol concentration, and acid concentration, bilinear effect of ethanol concentration and temperature, and quadratic effect of ethanol concentration. The predictive model can be generated from Table 5 in a similar way as done in the case of ethanol + acetic acid.

From the values of the parameter estimate of the linear factors contributing to the predictive model, it can be observed that ethanol concentration is the most significant linear factor

followed by ethanol concentration and HCl concentration. Based on the signs of the parameter estimates, with the increase in all of these factors the final response decreases, which can be confirmed by observing Figure 4b and c. The slope of the curves in the case of variation of  $\epsilon'$  with ethanol concentration is higher than in the case of variation with respect to temperature. The slope of the curve for the variation of  $\epsilon'$  with respect to HCl concentration is the least, and HCl does not contribute in any interaction effect. However, ethanol and temperature have a positive interaction effect; hence the  $\epsilon'$  measured at the point on the response surface curve presenting the highest ethanol concentration and temperature level is higher than the point with highest ethanol concentration and lowest temperature level (Figure 4a and b). As interpreted from the negative parameter estimate of the quadratic term  $\text{EtOH}^2$  (Table 5), there is a point of maximum for the response in the range of ethanol concentration used for the study.

Hence from regression analysis of combinations of both acids with ethanol, it can be interpreted that the  $\epsilon'$  will vary with temperature and the composition of the mixture solutions subjected to microwave radiations. The RSM gave us a broader idea about the effects of different factors and the pattern of variation of  $\epsilon'$  with different factors considered during the study.





**Figure 5.** Response surface plot representing effect of (a) HCl concentration ( $\text{mol}\cdot\text{dm}^{-3}$ , HCl) and temperature levels ( $^{\circ}\text{C}$ , temp) on dielectric loss of various HCl and ethanol concentration combinations at various temperature levels and (b) HCl concentration ( $\text{mol}\cdot\text{dm}^{-3}$ , HCl) and ethanol concentration ( $\text{cm}^3/100\text{ cm}^3$ , EtOH) on dielectric loss of various HCl and ethanol concentration combinations at various temperature levels.

**Dielectric Loss.** According to Table 5, ethanol concentration and temperature are the significant factors affecting the  $\epsilon''$ . The factors having linear effects on  $\epsilon''$  include ethanol concentration and temperature, and the only factor having quadratic effect is ethanol concentration, while there was no bilinear effect which can be used to generate the corresponding predictive model as discussed before.

According to the value of parameter estimates (Table 5), temperature is the most significant linear factor which is evident from the higher numerical value of parameter estimate for temperature than ethanol concentration. The negative sign of the parameter estimate for  $\text{EtOH}^2$  represents point of maximum of the response for the range of concentration of EtOH studied. Hence, the  $\epsilon''$  first increases with increase in concentration of ethanol, reaches a maximum, and then starts decreasing. In case of acetic acid, even at higher concentration, there is no significant effect of acid concentration.

In case of HCl, among the linear factors, temperature is the most significant factor followed by HCl and last ethanol concentration (Table 5). According to the sign of parameter estimates observed for  $\epsilon''$ , it can be interpreted that  $\epsilon''$  of ethanol with HCl combinations decreases with increase in temperature and ethanol concentration; however it increases with increase in HCl concentration. Among the bilinear terms, interactive terms of HCl with ethanol concentration and temperature are the significant terms. The negative value of parameter estimate for  $\text{EtOH}^2$  represents the increase in  $\epsilon''$  with an increase in the concentration of ethanol and a decrease with a further increase in ethanol concentration after achieving a peak value of  $\epsilon''$ .

From Figure 5a, it can be observed that, as temperature and HCl have a positive interactive parameter estimate, at the extreme points (lowest and highest factors' level combinations) the values are neither highest nor lowest and represent a mixed effect. In Figure 5b, it can be observed that, with an increase in ethanol concentration, there was an increase in  $\epsilon''$ ; however, at lowest HCl concentration and highest ethanol concentration the  $\epsilon''$  was lower than the maximum because of the declining effect of HCl with a decrease in its concentration and the significant negative bilinear effect of HCl and ethanol concentration.

In this section it can be observed that a higher concentration HCl (which is a strong acid) has a significant effect; however acetic acid does not have a significant effect on  $\epsilon''$  even at higher

concentration, which might be due to the fact that acetic acid is a weak acid. Hence the  $\epsilon''$  of a mixture solution at 2.45 GHz is affected by temperature and composition of mixture.

**Dissipation Factor.** Dissipation factor is determined by both the values of  $\epsilon'$  and  $\epsilon''$ . At 2.45 GHz, ethanol + acid solution mixtures had dissipation factors higher than 0.1, which is the  $\tan \delta$  of water; hence it can be interpreted that all of the combinations studied in this experimental setup can be potentially used as heating medium subjected to a microwave environment.<sup>8</sup>

Acetic acid + ethanol combinations are affected by the linear (temperature and ethanol concentration), bilinear (temperature-ethanol concentration), and quadratic (ethanol concentration<sup>2</sup>) effects of factors. According to the values of the parameter estimates in Table 5, it can be observed that, with an increase in temperature, the  $\tan \delta$  decreases and increases with an increase in ethanol concentration; hence the heating rate is also predicted to decrease with the decrease in ethanol concentration at higher temperature levels. However there is a negative bilinear effect of temperature with ethanol concentration. Hence, when the highest temperature will be applied combined with highest concentration of ethanol, it would give a comparatively lower value of  $\tan \delta$  due to the significant negative effect of temperature even if ethanol concentration is high (which has a positive effect). There is also a point of minimum for the response in terms of ethanol concentration as ethanol's quadratic effect has a positive parameter estimate.

In case of HCl, all of the factors are significant; hence all of the factors will affect the heating rate. It can be predicted from the values and signs of regression coefficients of the linear terms that the  $\tan \delta$  decreases with an increase in temperature and increases with an increase in ethanol concentration and HCl concentration. There is a negative interaction effect between temperature level and ethanol concentration; hence at lowest temperature and lowest ethanol concentration due to the lowering effect of low ethanol concentration, the  $\tan \delta$  can be found to be much lower than at the point with highest ethanol concentration, while at lowest ethanol concentration there will be a very low rate of increase of  $\tan \delta$  with the decrease in temperature. However, there is a positive interaction effect between temperature and HCl, and they increase each other's effect; hence, the highest value of  $\tan \delta$  will be observed at the highest HCl concentration

and lowest temperature level for the range of factors considered in this study. Furthermore, from the positive value of the regression coefficient of quadratic term with ethanol concentration, it can be interpreted that there is a point of minimum for the response in the range of ethanol concentration analyzed for the study. In case of  $\tan \delta$  it can be observed that there is a significant effect of HCl, which implies that, at a higher concentration of HCl, with the increase in concentration of HCl there will be an increase in the heating rate of the solution which can be attributed to the high polar nature.

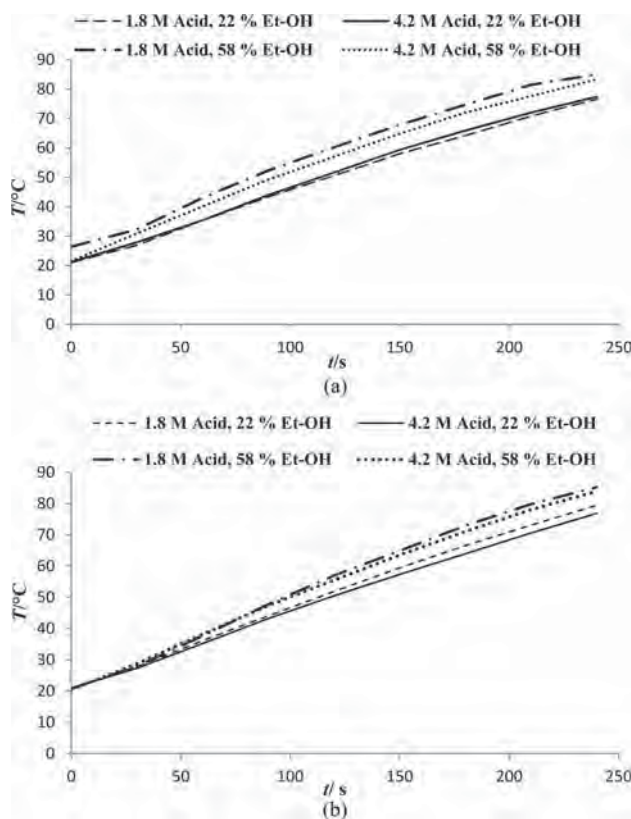
**Depth of Penetration.** Depth of penetration can also be used in the interpretation of heating rate. Generally it has been observed that for cases with higher  $D_p$ , the microwave heating rate is lower, which can be explained by the fact that as the  $D_p$  increases, the distance over which 63% of the microwave energy is dissipated increases,<sup>9</sup> hence the relative efficiency of the microwave heating method over the distance decreases.

In case of acetic acid, the ethanol concentration and temperature were found to be the significantly contributing factors. Linear effects (temperature and ethanol concentration), bilinear or interaction effect (temp·EtOH), and quadratic effect (ethanol-ethanol) were found to be significantly affecting the  $D_p$  in the case of combinations with acetic acid + ethanol. From the value and sign of the parameter estimates it can be interpreted that, with an increase in temperature, the  $D_p$  increases; however, with an increase in ethanol concentration, the  $D_p$  decreases. Hence the microwave heating rate can be interpreted to be decreasing with the increase in temperature and increasing with an increase in ethanol concentration. This observation is in agreement with the discussion in the previous sections. The acetic acid concentration does not contribute significantly to the  $D_p$  of the combination with acetic acid + ethanol, which might be related to the lower acidic strength of the acids. There is a negative interaction between the factors temperature and ethanol concentration, which implies that they decrease each other's effect on the  $D_p$ . The positive value of the quadratic effect of ethanol demonstrates that there is a point of minimum for the response in the ethanol concentration range used for the study.

In case of combinations with HCl + ethanol concentrations, temperature, ethanol, and HCl concentration contribute as significant linear effects; the interactive effect of HCl individually with temperature and ethanol concentration contributes as significant bilinear effects, and HCl contributes as the significant quadratic effect. Based on the values of parameter estimates, it can be interpreted that with an increase in temperature, and decrease in ethanol and HCl concentration,  $D_p$  for combinations with ethanol + HCl increases. Hence the microwave heating rate decreases with decrease in the concentration of ethanol and HCl and increase in temperature. Similar observations were discussed in the previous sections of this study, where the heating rate was interpreted to increase with the concentration of ethanol and HCl and decrease with the decrease in the concentration of ethanol and HCl. However, because of the negative interaction effect of temperature with HCl, it can be observed that these factors negatively influence each other's effect; hence at the highest level of HCl concentration the rate of increase of  $D_p$  with increasing temperature will be very low. Also it can be observed that, due to the positive quadratic effect of HCl, there is a point of minimum of the response for the concentration range of HCl used for the study, before which the  $D_p$  decreases with the increase in HCl concentration and after which the  $D_p$  increases with the increase in HCl concentration. Also there is a positive

interaction effect of ethanol and HCl which implies that they enhance each other's effect.

**Temperature Variation of Alcohol + Acid Combinations in the Microwave Environment.** Higher temperature levels were observed for higher concentrations of ethanol, and the divergence between the temperature levels attained for different concentrations of ethanol at a particular time interval increased with increase in the time of application of microwave. In case of acetic acid, the series 1 and 2 in Figure 6a have a similar



**Figure 6.** Graph showing the variation in temperature ( $T/^\circ\text{C}$ ) of various combinations of ethanol concentrations and acid (HCl and acetic acid) concentrations (molarity ( $M$ ),  $\text{mol}\cdot\text{dm}^{-3}$ ) while being heated in microwave environment (at  $71.05\text{ W}$ ) for time ( $t/\text{s}$ ) 240 s. Part a represents combinations with acetic acid and ethanol, and part b represents combinations with HCl and ethanol. Series 1 is the solution with  $1.8\text{ mol}\cdot\text{dm}^{-3}$  acid and  $22\%$  v/v ( $\text{cm}^3/100\text{ cm}^3$ ) ethanol; series 2 is the solution with  $4.2\text{ mol}\cdot\text{dm}^{-3}$  acid and  $22\%$  v/v ( $\text{cm}^3/100\text{ cm}^3$ ) ethanol; series 3 is the solution with  $1.8\text{ mol}\cdot\text{dm}^{-3}$  acid and  $58\%$  v/v ( $\text{cm}^3/100\text{ cm}^3$ ) ethanol; series 4 is the solution with  $4.2\text{ mol}\cdot\text{dm}^{-3}$  acid and  $58\%$  v/v ( $\text{cm}^3/100\text{ cm}^3$ ) ethanol.

temperature level with increasing time; also in case of series 3 and 4, which were initially at different temperatures, with an increase in time, the combinations achieved the same highest temperature level at a particular time interval of application of microwave. In case of combinations with HCl + ethanol and in cases with increased ethanol concentration, the difference between the temperature of solution achieved at the particular time interval of microwave application increased with the increase in the time of application of microwave. However, only in the case of HCl (and not in the case of acetic acid), the plots for solutions with the same ethanol concentration with different concentrations of HCl had moved apart with the increase in time of application. Hence,

HCl can be interpreted to be significantly affecting the temperature of solution in a microwave environment, along with the ethanol concentration. This observation is similar to the interpretation from the earlier discussion that, with an increase in concentration of HCl and ethanol, the heating rate increases.

## CONCLUSION

At higher concentrations of acetic acid, temperature and ethanol concentration were the significant factors affecting the dielectric properties,  $\tan \delta$  and  $D_p$ , of the combinations with ethanol + acetic acid. However, in the case of combinations with higher concentration of HCl with ethanol, HCl was also significantly contributing to the various responses along with the other factors. The dielectric properties measured were efficiently used for interpreting the microwave heating rate and gave similar conclusive observations regarding the different factors. It was interpreted from the results that, with increasing concentration of ethanol and HCl acid and lower temperature levels, the microwave heating rate is increased. However, due to the limitation of highest temperature achieved during microwave heating in this case, the decrease of heating rate with increasing temperature was not observed beyond 90 °C (363.15 K), which is expected to be more significantly observed at higher temperature levels than 90 °C (363.15 K). Overall, the measurement of dielectric properties and application of CCD-RSM to analyze the data gave us sufficient information regarding the behavior of solution mixtures subjected to microwave and helped in predicting their heating rate. Further studies with microwave reactors equipped to achieve a higher temperature will help in proving the effect of higher temperatures on novel mixtures of different solvent components.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: valerie.orsat@mcgill.ca.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank Dr. G. S. V. Raghavan (James McGill Professor, Bioresource Engineering Department, Macdonald Campus, McGill University, Canada) for providing us the required equipment and facilities for the experiments. We also thank Mr. Yvan Gariépy (Research Associate, Bioresource Engineering Department, Macdonald Campus, McGill University, Canada) for his expertise and support in conducting the experiments.

## REFERENCES

- (1) Alfaro, M. J.; Bélanger, J. M. R.; Padilla, F. C.; Jocelyn Pare, J. Influence of Solvent, Matrix Dielectric Properties, and Applied Power on the Liquid-Phase Microwave-Assisted Processes (MAP (TM)) 1 Extraction of Ginger (*Zingiber officinale*). *Food Res. Int.* **2003**, *36*, 499–504.
- (2) Jain, T.; Jain, V.; Pandey, R.; Vyas, A.; Shukla, S. S. Microwave Assisted Extraction for Phytoconstituents – An Overview. *Asian J. Res. Chem.* **2009**, *2*, 19–25.
- (3) Vian, M. A.; Fabiano-Tixier, A. S.; Elmaataoui, M.; Dangles, O.; Chemat, F. A Remarkable Influence of Microwave Extraction: Enhancement of Antioxidant Activity of Extracted Onion Varieties. *Food Chem.* **2011**, *127*, 1472–1480.
- (4) Cacace, J. E.; Mazza, G. Optimization of Extraction of Anthocyanins from Black Currants with Aqueous Ethanol. *J. Food Sci.* **2003**, *68*, 240–248.

(5) *Handbook of Food Analytical Chemistry-Pigments, Colorants, Flavors, Texture, and Bioactive Food Components*; Giusti, M. M., Wrolstad, R. E., Acree, T. E., Decker, E. A., Penner, M. H., Reid, D. S., Schwartz, S. J., Shoemaker, C. F., Smith, D., Eds.; John Wiley and Sons, Inc.: Hoboken, NJ, 2005.

(6) Sun, Y.; Liao, X.; Wang, Z.; Hu, X.; Chen, F. Optimization of Microwave-Assisted Extraction of Anthocyanins in Red Raspberries and Identification of Anthocyanin of Extracts Using High-Performance Liquid Chromatography–Mass Spectrometry. *Eur. Food Res. Technol.* **2007**, *225*, 511–523.

(7) Krishnaswamy, K.; Orsat, V.; Gariépy, Y.; Thangavel, K. Optimization of Microwave-Assisted Extraction of Phenolic Antioxidants from Grape Seeds (*Vitis vinifera*). *Food Biopro. Technol.* **2013**, *6* (2), 441–455.

(8) Gabriel, C.; Gabriel, S.; Grant, E. H.; Halstead, B. S. J.; Mingos, D. M. P. Dielectric Parameters Relevant to Microwave Dielectric Heating. *Chem. Soc. Rev.* **1998**, *27*, 213–224.

(9) Liao, X.; Raghavan, G.; Yaylayan, V. Dielectric Properties of Alcohols (C1-C5) at 2450 and 915 MHz. *J. Mol. Liq.* **2001**, *94*, 51–60.

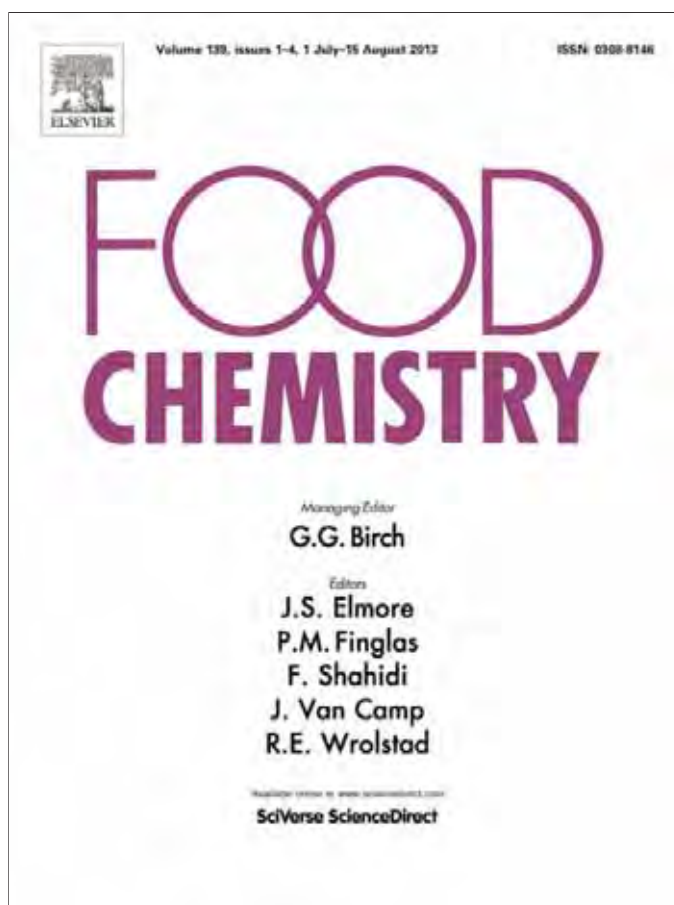
(10) Yang, L. J.; Yang, X. Q.; Huang, K. M.; Shang, H.; Jia, G. Z. Experimental and Theoretic Study of the Dielectric Properties of Ethanol+ Methanol Mixtures. *J. Solution Chem.* **2010**, *39*, 473–481.

(11) Mohsen-Nia, M.; Amiri, H.; Jazi, B. Dielectric Constants of Water, Methanol, Ethanol, Butanol and Acetone: Measurement and Computational Study. *J. Solution Chem.* **2010**, *39*, 1–8.

(12) Myers, R. H.; Montgomery, D. C.; Anderson-Cook, C. M. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*; John Wiley & Sons, Inc.: New York, 2009.

(13) Nemes, S. M. Practical Methods for Lignans Quantification. Ph.D. Thesis Bioresource Engineering, McGill University, Montreal, Canada, 2012.

Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

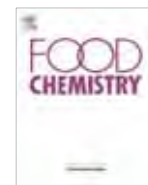
<http://www.elsevier.com/authorsrights>





Contents lists available at SciVerse ScienceDirect

Food Chemistry

journal homepage: [www.elsevier.com/locate/foodchem](http://www.elsevier.com/locate/foodchem)

## Fatty acid profiling of the seed oils of some varieties of field peas (*Pisum sativum*) by RP-LC/ESI-MS/MS: Towards the development of an oilseed pea

Manuel Ivan Villalobos Solis<sup>a</sup>, Anil Patel<sup>a</sup>, Valérie Orsat<sup>a</sup>, Jaswinder Singh<sup>b</sup>, Mark Lefsrud<sup>a,\*</sup>

<sup>a</sup> Bioresource Engineering Department, McGill University, Macdonald Campus, 21,111 Lakeshore Road, Ste-Anne-de-Bellevue, Quebec, Canada H9X 3V9

<sup>b</sup> Plant Science Department, McGill University, Macdonald Campus, 21,111 Lakeshore Road, Ste-Anne-de-Bellevue, Quebec, Canada H9X 3V9

### ARTICLE INFO

#### Article history:

Received 7 August 2012

Received in revised form 30 November 2012

Accepted 21 December 2012

Available online 16 January 2013

#### Keywords:

*Pisum sativum*

Seed oil

RP-LC/ESI-MS/MS

Fatty acids

### ABSTRACT

Reversed-phase liquid chromatography coupled to negative-ion electrospray tandem mass spectrometry (RP-LC/ESI-MS/MS) was used to study the fatty acid profile from the oil of harvested field pea (*Pisum sativum*) varieties as part of a research project to develop this legume as a commercial oilseed for Canada. The seed oils from pea samples contained palmitic and stearic acids as major saturated fatty acids. Oleic, linoleic and linolenic acids were the major unsaturated fatty acids found. Small percentages of other long chain fatty acids were also detected. This profile suggests that the species of field pea investigated might have the potential to be used as raw materials to develop a future new oilseed crop for the food industry. Fatty acid extracts did not require further manipulation before final analysis by RP-LC/ESI-MS/MS, indicating the utility and relative simplicity of this technique for future screening studies.

© 2013 Elsevier Ltd. All rights reserved.

### 1. Introduction

The field pea (*Pisum sativum* L.) also known as common pea, dry pea, green pea, yellow pea and garden pea (Ratnayake, Hoover, Shahidi, Perera, & Jane, 2001) is one of the major food legumes in the world and ranks fourth in the world production next to soybeans, peanuts and dry beans (Yoshida, Tomiyama, Tanaka, & Mizushima, 2007). Canada is the world largest producer and exporter of field peas. In 2010 of the 2.86 metric tonnes produced in Canada, approximately 65% was in Saskatchewan while Alberta and Manitoba produced 33% and 2% respectively (Pulse Canada. Canadian pulse industry: Situation, 2012).

This crop has an optimum growth temperature range of 12–18 °C and a cold tolerance down to –7 °C. Field pea protein content has been calculated at 23% (Thiessen, 2004; Wang & Daun, 2004) and as high as 25% (Singh, Kaur, Rana, & Sharma, 2010) and its carbohydrate content falls in the 50–60% range (Nikolopoulou, Grigorakis, Stasini, Alexis, & Iliadis, 2007). It is underutilised as whole or split in soups and stews, as hulls in high fibre bread, in the production of adhesives and carbonless paper (pea starch) and as an alternative protein source to soybean (*Glycine max*) and canola (*Brassica napus*) meal (Ratnayake et al., 2001). Published literature has shown low average oil content in field peas, ranging from 1.5% to 3.7% (Welch & Wynne Griffiths, 1984; Yoshida

et al., 2007) suggesting that current varieties have limited use as an oil crop.

A recent study conducted at the Bioresource Engineering Department of McGill University screened the lipid content of 198 varieties of field pea and found 30 candidates with mean contents in the range of 1.15–6.31%. With further research, these oil yields could allow the development of the field pea as an oilseed with a commercial lipid content greater than 20%, and potentially compete with the primary oilseeds grown in Canada, namely, canola (with >35% oil) and soybean (>20% oil). To achieve this long-term objective, the development of a new variety of field pea requires two major steps: screening for lipid content and its improvement through breeding.

Benefits of a new oilseed feedstock for Canada include cold weather tolerance with similar annual grain yields as soybean, nitrogen fixation (with nitrogen storage of about 65–70%) that could allow a reduction in the costs of fertilisers, and a source of high quality protein.

A field pea variety rich in poly-unsaturated fatty acids ( $\omega$ -3,  $\omega$ -6), could become a dietary source contributing to cardiovascular health, physiological functions and maintenance (Ryan, Galvin, O'Connor, Maguire, & O'Brien, 2007). Furthermore, field pea as an alternative oilseed holds potential as a bio-oil resource for the production of oleochemicals derived from it. One example is commercial biodiesel. By 2011 the government of Canada targeted that 2% of the diesel fuel and heating oil needed to be from renewable sources, where an oilseed pea variety could be developed to compete with canola and soybean. Other industrial applications include lubricants produced from high oxidative stability oils

\* Corresponding author. Tel.: +1 514 398 7967; fax: +1 514 398 8387.

E-mail addresses: [manuel.villalobosolis@mail.mcgill.ca](mailto:manuel.villalobosolis@mail.mcgill.ca) (M.I. Villalobos Solis), [anil.patel@mail.mcgill.ca](mailto:anil.patel@mail.mcgill.ca) (A. Patel), [valerie.orsat@mcgill.ca](mailto:valerie.orsat@mcgill.ca) (V. Orsat), [jaswinder.singh@mcgill.ca](mailto:jaswinder.singh@mcgill.ca) (J. Singh), [mark.lefsrud@mcgill.ca](mailto:mark.lefsrud@mcgill.ca) (M. Lefsrud).



(e.g. high oleic acid content) or coating applications such as paints, inks and varnishes from low oxidative stability oils (e.g. high linolenic acid content) (Cahoon, 2003).

Fundamental to screening varieties of field peas for high oil content is an appropriate method for the characterisation of fatty acids. Fatty acid profiling can give information about suitable candidates for breeding and selection experiments, classification of newly produced varieties and information on appropriate strains for different applications, be it oleochemicals production or human consumption.

However, the analysis of lipids is quite complicated because of the high degree of complexity and the heterogeneity of their components. Especially from vegetable samples, the number of lipid molecules is higher than that of animal or prokaryotic organisms hence introducing more degree of diversity (Seiwert, Giavalisco, & Willmitzer, 2009). At present, there is no single analytical tool capable of simultaneously identifying and quantifying all these lipid species without combining various technologies.

Gas chromatography coupled to flame ionisation (GC–FID) has been the method of choice for fatty acid screening studies but it requires sample derivatisation which is time-consuming and may compromise the stability of the molecules with undesirable alterations such as isomerisation, oxidation and thermal degradations (Nishiyama-Naruke, Souza, Carnelos, & Curi, 1998); for these reasons, more practical alternatives have been explored in recent years which include the use of liquid chromatography coupled to mass spectrometry (LC–MS). This technique also allows fatty acids with the same equivalent carbon number (ECN) to be discriminated by their masses or fragments (Byrdwell, 2005, chap. 5), a task that is relatively more difficult to achieve with traditional two dimensional detectors like refractive index, UV/Vis and evaporative light scattering detector (ELSD).

According to Lima and Abdalla (2002), some advantages that electrospray ionisation in the LC–MS analysis of fatty acids offers are improved separation of these molecules from the complex biological matrices; the possibility of using high flows (100–1000  $\mu\text{L min}^{-1}$ ) usually necessary in HPLC separations without the loss of sensitivity; and the omission of chemical derivatisations, like esterification or transesterification prior to analysis, which in turns speeds the assay, reducing it to mainly two steps: saponification and extraction of fatty acids (Kurata, Yamaguchi, & Nagai, 2005).

By employing electrospray ionisation in the negative mode, ions of fatty acids can be monitored as deprotonated molecules ( $[M-H]^-$ ) with the aid of acetic acid, formic acid or ammonium acetate which are added to the mobile phases in order to facilitate the ionisation (Kerwin, Wiens, & Ericsson, 1996; Perret, Gentili, Marchese, Sergi, & Caporossi, 2004). By conducting tandem mass spectrometry, it is possible to determine branched points or double bonds in the fatty acid chains, providing a more accurate detection method (Lima & Abdalla, 2002).

The aim of the present work was to develop an analytical method for the identification of fatty acids in pea oil samples based on reverse phase high performance liquid chromatography interfaced to negative ion electrospray tandem mass spectrometry (RP-LC/ESI-MS/MS) in order to help in the long term development of field pea as an oilseed.

Oil extraction was performed by a modified version of the hexane: isopropanol method proposed by Ryan, Galvin, O'Connor, Maguire, and O'Brien (2007) which according to previous experiments in our laboratory had the highest yields of pea oil (unpublished data). A mild alkaline hydrolysis procedure at low temperature was also employed in order to release the fatty acid molecules from complex lipid structures.

The analytical methodology proposed here may also prove useful for the separation and analysis of fatty acids from other seed oil sources as it was also tested for samples of soybean and canola.

## 2. Materials and methods

### 2.1. Biological samples

Six varieties of field pea (*P. sativum*) catalogued as 29579, 29595, 29610, 45760, Big Pea and Galena were analysed in this study (Plant Gene Resources of Canada, Saskatoon, SK; U.S. Department of Agriculture, Pullman, WA). Seeds were harvested in the year 2010 from either Lefsrud's farm (Viking, AB, Canada) and/or Macdonald Campus of McGill University (Ste-Anne-de-Bellevue, QC, Canada). Pods of the samples were previously dried in an oven at 60 °C for 48 h with no further treatment. Seeds from cultivars of soybean (*G. max*), OAC Champion variety, and canola (*B. napus*), Roper variety, were also screened for fatty acid profile and these were grown and harvested at Lefsrud's farm (Viking, AB, Canada) and Agrocentre Belcan (Sainte-Marthe, QC, Canada) respectively.

### 2.2. Chemicals and reagents

For sample preparation the following analytical grade solvents and reagents were used: 2-propanol LC–MS Chromasolv<sup>®</sup> was purchased from Sigma–Aldrich (St. Louis, MO, USA); *n*-hexane 95% optima; chloroform HPLC-grade 99.8; petroleum ether certified ACS; hydrochloric acid reagent A.C.S.; sodium sulphate anhydrous granular, 10–60 Mesh, Certified ACS and potassium hydroxide, extra pure, ca. 85% flakes, were all obtained from Fisher Scientific (Fair Lawn, NJ, USA). For LC–MS analysis, acetic acid was acquired from Sigma–Aldrich (St. Louis, MO, USA); Optima<sup>®</sup> LC–MS grade acetonitrile, methanol and water were purchased from Fisher Scientific (Fair Lawn, NJ, USA). All solvents were used without further purification.

The saturated fatty acid standards, myristic ( $C_{14:0}$ ), palmitic ( $C_{16:0}$ ), stearic ( $C_{18:0}$ ), arachidic ( $C_{20:0}$ ), behenic ( $C_{22:0}$ ) and lignoceric ( $C_{24:0}$ ); as well as the mono- and polyunsaturated standards, palmitoleic ( $C_{16:1}$ ), oleic ( $C_{18:1}$ ), linoleic ( $C_{18:2}$ ),  $\alpha$ - and  $\gamma$ -linolenic ( $C_{18:3}$ ), gadoleic ( $C_{20:1}$ ), eicosadienoic ( $C_{20:2}$ ), erucic ( $C_{22:1}$ ) and nervonic ( $C_{24:1}$ ) were purchased from Nu-Check Prep (Elysian, MN, USA). Individual stock solutions containing 17  $\mu\text{g}/\mu\text{L}$  of each standard were prepared by dissolving the analytes in HPLC-grade chloroform; composite working standard solutions were prepared by mixing adequate volumes of diluted stock solutions and adding acetonitrile as needed. All solutions were flushed with nitrogen every time after use and stored at –80 °C.

### 2.3. Sample preparation

#### 2.3.1. Lipid extraction and oil determination

Oil extraction from the different biological samples was achieved by a modified version of the method described by Ryan, Galvin, O'Connor, Maguire, and O'Brien (2007). Eight grams of ground samples were placed in 50 ml Teflon-lined screw-capped glass centrifuge tubes of known mass in three biological replicates. A fourth tube with the same amount of ground sample was prepared and used as a control in order to measure moisture content.

Oil was extracted with 12 ml of hexane–isopropanol (3:2, v/v) at room temperature and constant stirring for 1 h. The samples were then vortex for 30 s followed by centrifugation at 3000g for 10 min. The solvent layers were collected in 50 ml glass centrifuge tubes. The pellets were washed two more times with 10 ml of the organic solvent, each time vortexing for 30 s, centrifuging at 3000g for 10 min and recovering the hexane: isopropanol layers. The washed pellets were dried under nitrogen for up to 30 min at 70 °C until the remaining solvent was completely evaporated. The tubes were un-capped and placed in the oven for 24 h at 95 °C along with the control tubes for each sample. After the drying

was over and the tubes were allowed to reach room temperature their final masses were measured and recorded. The difference between the initial and final mass of the control tube, which represents the moisture loss during the drying period, was subtracted from the difference in mass of each sample tube in order to calculate the oil content percentage of the samples.

The recovered hexane–isopropanol extracts were centrifuged at 3000g for 10 min and the supernatant transferred to a new known mass glass centrifuge tube. The solvent was nitrogen evaporated under 50 °C and the pure oil was redissolved in chloroform for storage and further sample preparation.

### 2.3.2. Fatty acids extraction by saponification

After solvent evaporation, approximately 100 mg of extracted oil from each sample were treated with 5 ml of methanolic KOH (0.5 M) at 55 °C for 20 min. When the saponification reaction was ready, 2.5 ml of concentrated HCl were added. The mixture was then transferred to a 30 ml separatory funnel and the fatty acids extracted twice with 5 ml of petroleum ether.

### 2.3.3. Removal of water and non-lipid contents

The combined petroleum ether extracts from the saponification reaction were transferred into glassware funnels containing 15 g of anhydrous sodium sulphate previously rinsed with 5 ml of petroleum ether. The filtrates were collected in 50 ml glass centrifuge tubes. Additional washes of 2 ml petroleum ether were done and collected in the same tubes. Nitrogen was gently blown through the combined washes at 40 °C. When most of the solvent was evaporated from the solutions, the samples were re-dissolved with 10 ml of HPLC grade chloroform, flushed with nitrogen and stored at –80 °C. From these stock solutions 3 ml were diluted each time with 7 ml of LC–MS grade acetonitrile for further RP-LC/ESI-MS/MS analysis.

## 2.4. RP-LC/ESI-MS/MS analysis

### 2.4.1. RP-LC system

The HPLC system was a Surveyor Plus™ LC equipped with an analytical pump, an autosampler with a 25 µL sample loop and a UV/Vis detector (Thermo Electron, San Jose, CA, USA). A reverse phase Hypersil Gold C18 Column® with a particle size of 5 µm (100 × 4.6 mm, i.d. × length) (Thermo Fisher Scientific, Waltham, MA, USA) was used to analyse the fatty acid fractions collected. A ternary solvent gradient elution method consisting of 25% acetonitrile–75% water (mobile phase A), 100% acetonitrile (mobile phase B) and 100% methanol (mobile phase C) all containing 0.12% acetic acid was programmed to separate the molecules. The mobile phases changed from 46.00% A, 54.00% B and 0.00% C at time zero to 12.00% A, 85.00% B and 3.00% C in 17 min; from the last percentages to 12.00% A, 0.00% B and 88.00% C in 9 min and then to 0.00% A, 0.00% B and 100.00% C in 1 min. This gradient remained constant for 7 min, then returned to the initial composition in 1 min and maintained for 10 additional minutes. The flow rate employed at all times was 1 ml min<sup>-1</sup>. The column was kept at a constant temperature of 35 °C. 10 µL of each biological replicate were injected three times (technical replicates) into the column for analysis. The effluent from the LC system was split 1:10 before introduction to the ESI source.

### 2.4.2. Tandem mass spectrometry

A LTQ-XL 2D linear ion trap mass spectrometer (Thermo Electron, San Jose, CA, USA) equipped with an ESI source was coupled on-line to the RP-LC system described before. The spectrometer was operated in the negative ion (NI) mode. The source voltage was 3.50 kV. Capillary temperature was maintained at 350.00 °C. The settings for the sheath and auxiliary gas flows were respec-

tively 30.00 and 5.00 instrument units. Data acquisition and processing were performed using Xcalibur® version 2.0.7 software (Thermo Electron, San Jose, CA, USA).

A data dependent tandem mass spectrometric (MS/MS) experiment consisting of two scan events was programmed on the data system in order to identify the fatty acid molecules in the extracts. The full scan mode was chosen as the acquisition mode. The mass spectra were obtained between *m/z* 200 and 800. During MS/MS, helium was used as the collision gas for collision induced dissociation (CID). The collision energy was set at 35.00 eV and the isolation width set at 2.0 u. The system was calibrated by infusion of a normal mass range calibration solution containing caffeine, L-methionyl–arginyl–phenylalanyl–alanine acetate salt (MRFA) and Ultramark 1621 at a flow rate of 5 µL/min.

## 2.5. Data evaluation

Fatty acids were identified based on elution patterns from standard working solutions and characteristic parent peaks detected in the mass spectra of each compound. Structure elucidation and confirmation were conducted taking as reference the MS/MS spectra of each standard. The fatty acid composition was calculated as a relative percentage of the total area of all peaks.

## 3. Results and discussion

### 3.1. Fatty acid composition of pea oil

The obtained yields of pea oil (from 1.50% for variety 29579 to 2.03% for variety Galena), soybean oil (15.32%) and canola oil (36.10%) are in close agreement with what literature has reported for these crops (Hammond, Johnson, Su, Wang, & White, 2005; Ryan, Galvin, O'Connor, Maguire, & O'Brien, 2007; Shahidi, 1990, chap. 6; Wang & Daun, 2004; Welch & Wynne Griffiths, 1984; Yoshida et al., 2007). According to Yoshida et al. (2007) predominant lipid components of field pea oil are phospholipids (PLs) and triacylglycerols (TAGs). These authors analysed four varieties of field pea and found that between both lipid classes the amount of PLs (55.20–61.30 wt.%) was higher than that of TAGs (31.20–40.30 wt.%) mainly because peas are not oilseeds, but typical vegetable seeds. The same researchers also reported minor quantities of free fatty acids (1.3–2.7 wt.%), 1,2-diacylglycerols (1.0–2.2 wt.%), 1,3-diacylglycerols (1.0–1.8 wt.%), steryl esters (0.8–2.4 wt.%) and hydrocarbons (0.5–0.9 wt.%) for the same varieties.

The previous data suggests that the bulk of the fatty acid fraction of pea oil is bound to other molecules and this is the reason of why a saponification procedure was employed here.

The fatty acid profile of the different varieties of field pea, canola and soybean samples analysed by RP-LC/ESI-MS/MS and their relative percentage in oil are presented in Table 1. Linoleic acid (C<sub>18:2</sub>) was the major component in field peas and soybean oil; whereas oleic acid (C<sub>18:1</sub>) occupied the highest percentage in canola oil; all these results agree with published data (El-Saied, Amer, & Gabran, 1981; Hammond et al., 2005; Murcia & Rincon, 1992; Nishiyama-Naruke et al., 1998; Ryan, Galvin, O'Connor, Maguire, & O'Brien, 2007; Shahidi, 1990, chap. 6; Welch & Wynne Griffiths, 1984; Yoshida et al., 2007).

The levels of saturated fatty acids (SFAs) in all samples ranged from 10.83% for Canola Roper to 20.13% for pea Galena variety. Total mono-unsaturated fatty acids (MUFAs) ranged from 26.60% for pea variety 45760 to 58.15% for Canola Roper. Poly-unsaturated fatty acids (PUFAs) ranged from 30.92% for Canola Roper to 56.80% for pea variety 29610. The most abundant PUFA in all samples was linoleic acid, MUFA was oleic acid and SFA was palmitic acid (C<sub>16:0</sub>). Unknown molecules were also seen in the different

**Table 1**

Fatty acid composition of the oilseeds analysed. The peak area of each component is reported as a percent of the total area of all peaks. Values are the means of three biological replicates  $\pm$  standard deviation.

Fatty acids or <i>m/z</i> values detected	Relative peak area (% of fatty acids in extracted oil)							
	29579	29595	Big Pea	29610	45760	Galena	OAC champion	Roper
227 <sup>a</sup>	0.09 $\pm$ 0.02	ND	ND	ND	ND	ND	ND	ND
C <sub>16:0</sub>	9.01 $\pm$ 0.11	7.54 $\pm$ 0.52	7.63 $\pm$ 0.64	7.17 $\pm$ 0.23	8.91 $\pm$ 0.52	7.20 $\pm$ 0.32	8.77 $\pm$ 0.18	3.67 $\pm$ 0.30
C <sub>18:0</sub>	5.12 $\pm$ 0.24	5.48 $\pm$ 0.31	5.02 $\pm$ 0.15	4.89 $\pm$ 0.14	6.27 $\pm$ 0.75	5.94 $\pm$ 0.49	7.31 $\pm$ 0.73	3.44 $\pm$ 0.25
C <sub>18:1</sub>	26.9 $\pm$ 0.07	35.79 $\pm$ 1.35	27.87 $\pm$ 0.74	26.56 $\pm$ 1.94	25.52 $\pm$ 0.80	34.51 $\pm$ 0.91	27.39 $\pm$ 0.63	54.90 $\pm$ 3.16
C <sub>18:2</sub>	39.45 $\pm$ 1.34	35.95 $\pm$ 2.22	42.32 $\pm$ 3.55	44.78 $\pm$ 0.89	43.98 $\pm$ 2.22	35.85 $\pm$ 3.64	43.58 $\pm$ 1.33	21.38 $\pm$ 0.91
C <sub>18:3</sub>	14.01 $\pm$ 0.60	9.59 $\pm$ 0.72	11.25 $\pm$ 1.28	12.13 $\pm$ 0.58	9.89 $\pm$ 0.59	6.01 $\pm$ 0.26	8.97 $\pm$ 0.64	9.54 $\pm$ 1.36
C <sub>20:0</sub>	1.29 $\pm$ 0.10	1.17 $\pm$ 0.17	1.16 $\pm$ 0.29	1.09 $\pm$ 0.19	1.30 $\pm$ 0.10	1.90 $\pm$ 0.18	0.85 $\pm$ 0.14	1.32 $\pm$ 0.10
C <sub>20:1</sub>	0.76 $\pm$ 0.09	0.84 $\pm$ 0.37	0.91 $\pm$ 0.14	0.72 $\pm$ 0.09	0.90 $\pm$ 0.12	1.50 $\pm$ 0.27	0.30 $\pm$ 0.03	2.76 $\pm$ 0.23
307 <sup>a</sup>	ND	ND	0.14 $\pm$ 0.01	ND	ND	ND	ND	ND
C <sub>22:0</sub>	0.57 $\pm$ 0.09	ND	0.46 $\pm$ 0.14	0.51 $\pm$ 0.34	0.51 $\pm$ 0.10	1.20 $\pm$ 0.35	1.67 $\pm$ 0.25	1.27 $\pm$ 0.21
C <sub>22:1</sub>	0.35 $\pm$ 0.04	0.50 $\pm$ 0.19	0.29 $\pm$ 0.09	0.25 $\pm$ 0.04	0.18 $\pm$ 0.02	0.37 $\pm$ 0.06	ND	ND
C <sub>24:0</sub>	1.78 $\pm$ 0.10	2.73 $\pm$ 0.70	1.90 $\pm$ 0.72	1.26 $\pm$ 0.35	1.55 $\pm$ 0.14	3.88 $\pm$ 1.19	1.07 $\pm$ 0.27	1.12 $\pm$ 0.19
C <sub>24:1</sub>	ND	ND	ND	ND	ND	ND	ND	0.50 $\pm$ 0.16
325 <sup>a</sup>	ND	ND	ND	ND	0.08 $\pm$ 0.01	0.15 $\pm$ 0.01	ND	ND
353 <sup>a</sup>	ND	ND	ND	ND	0.16 $\pm$ 0.04	0.35 $\pm$ 0.08	ND	ND
381 <sup>a</sup>	0.24 $\pm$ 0.03	ND	0.35 $\pm$ 0.14	0.26 $\pm$ 0.08	0.26 $\pm$ 0.04	0.55 $\pm$ 0.14	0.03 $\pm$ 0.01	ND
395 <sup>a</sup>	0.34 $\pm$ 0.02	0.42 $\pm$ 0.03	0.40 $\pm$ 0.09	0.24 $\pm$ 0.05	0.55 $\pm$ 0.03	0.57 $\pm$ 0.10	0.07 $\pm$ 0.01	0.10 $\pm$ 0.01
SFAs	17.86 $\pm$ 0.64	16.92 $\pm$ 1.47	16.17 $\pm$ 1.89	14.92 $\pm$ 0.88	18.54 $\pm$ 1.39	20.13 $\pm$ 2.40	19.67 $\pm$ 1.48	10.83 $\pm$ 0.90
MUFAs	28.10 $\pm$ 0.08	37.29 $\pm$ 0.86	29.08 $\pm$ 0.69	27.54 $\pm$ 1.86	26.60 $\pm$ 0.70	36.39 $\pm$ 0.96	27.69 $\pm$ 0.65	58.12 $\pm$ 2.85
PUFAs	53.46 $\pm$ 0.76	44.68 $\pm$ 1.59	53.51 $\pm$ 2.28	56.80 $\pm$ 1.17	53.87 $\pm$ 2.11	41.86 $\pm$ 3.42	52.55 $\pm$ 1.96	30.92 $\pm$ 2.23
Unknown	0.58 $\pm$ 0.05	0.42 $\pm$ 0.03	1.19 $\pm$ 0.33	0.63 $\pm$ 0.17	1.05 $\pm$ 0.07	1.62 $\pm$ 0.32	0.09 $\pm$ 0.02	0.10 $\pm$ 0.01

ND, not detected; SFAs, Saturated fatty acids; MUFAs, mono-unsaturated fatty acids; PUFAs, poly-unsaturated fatty acids.

<sup>a</sup> Undefined *m/z* value detected.

**Table 2**

Comparative table of the fatty acid composition of pea oil (%). Values are mean determinations.

Fatty acid	Mean values found (%)	Reported values (%)			
		El-Saied et al. (1981) <sup>a</sup>	Murcia and Rincon (1992) <sup>b</sup>	Wang and Daun (2004) <sup>a</sup>	Ryan, Galvin, O'Connor, Maguire, and O'Brien (2007) <sup>a</sup>
Palmitic (C <sub>16:0</sub> )	6.76	12.79	16.40	10.65	10.65
Stearic (C <sub>18:0</sub> )	2.86	2.41	15.20	3.29	3.29
Oleic (C <sub>18:1</sub> )	31.04	14.67	23.50	23.22	28.15
Linoleic (C <sub>18:2</sub> )	46.06	53.99	32.90	45.63	47.59
Linolenic (C <sub>18:3</sub> )	11.12	9.04	12.00	13.69	9.29
Arachidic (C <sub>20:0</sub> )	0.13	NR	NR	0.79	0.22
Gadoleic (C <sub>20:1</sub> )	0.25	NR	NR	0.62	0.21
Erucic (C <sub>22:1</sub> )	0.03	NR	NR	0.24	ND
Lignoceric (C <sub>24:0</sub> )	1.77	NR	NR	0.33	NR

NR – not reported, ND – not detected.

<sup>a</sup> Determined by GC–FID.

<sup>b</sup> Determined by GC–MS.

samples, but their relative percentages were low (less than 1.50%), a discussion about their possible identities is presented in the mass spectrometry analysis section of this paper.

Palmitic acid was detected in all field pea varieties but found at the highest level in 29579 (9.01%) than in other species. Stearic acid (C<sub>18:0</sub>) was higher in field pea seed from variety 45760 (6.27%) than in all others. Arachidic acid (C<sub>20:0</sub>) was present in all pea samples at relative similar percentages from 1.09% for variety 29610 to 1.90% for Galena. Behenic acid (C<sub>22:0</sub>) occurred in field pea samples at less than 1.50%, with Galena variety having the highest percentage (1.20%) and Big Pea variety the lowest (0.46%). According to Balogun and Fetuga (1985) oils with high levels of behenic acid may be difficult to digest by humans and animals, thus highlighting the value that pea oil could have for food purposes. Other SFA found in this study was lignoceric acid (C<sub>24:0</sub>) a long chain fatty acid that has not been reported extensively for field pea oil but has been found in other members of the Leguminosae family (Bağcı & Şahin, 2004).

The major MUFA found in all samples was oleic acid. Specifically for field peas, seeds from varieties 29595 (35.79%) and Galena (34.51%) had the highest levels of it. Other acids found and belonging to this category were gadoleic (C<sub>20:1</sub>) and erucic (C<sub>22:1</sub>) acids.

The first of them had the highest level, within varieties of field pea, in oil from variety Galena (1.50%), whereas erucic acid presented its highest level in variety 29595. Nervonic acid (C<sub>24:1</sub>) was not present in field peas neither in soybean, but canola had a level of 0.50%.

The isomer  $\alpha$ -linolenic acid (C<sub>18:3</sub>) was found in all samples but its levels were lower than oleic and linolenic acid as stated before. Studies by Hilditch and Williams (1964) suggested that linolenic acid is either absent or present in very small amounts in most Leguminosae seed oils; however, the percentages calculated here, resulted in the third largest in importance as opposed to the trace amounts (<0.05%) found of arachidic, gadoleic and lignoceric acids.

A comparative table (Table 2) of the fatty acid composition obtained in this study for pea oil to what other authors have found by gas chromatography–flame ionisation detector (GC–FID) or gas chromatography–mass spectrometry (GC–MS) is presented.

According to Murcia and Rincon (1992) fatty acid composition of pea oil is influenced by agro-climatic conditions as it is known that the enzymes in fatty acid biosynthesis depend on these factors. The reduced quantity of palmitic and linolenic acids found in peas in comparison to the levels of linoleic acid can be explained by the natural growth of these legumes as it has been observed

that, as the diameter of peas increases during growth period, the proportion of these fatty acids diminishes. Polyunsaturated fatty acids in pea oil are commonly found in seed oils, and they play a role in the initial oxidative processes whereby fatty acids are converted to carbohydrates, because of their ease of peroxidation (Holman, 1981).

Our results indicate that due to the content of PUFAs, pea oil can be advisable for human consumption. Oleic acid has been reported as effective in reducing cardiovascular risk; linoleic acid, which was the major component of pea oil, is considered an “essential” fatty acid for human growth, physiological functions and maintenance; and  $\alpha$ -linolenic acid is regarded as potentially functional in reducing cardiovascular risk through its particular omega-3 structure (Shahidi, 1990, chap. 6). The nutritional value of pea oil is also shared with soybean and canola as their fatty acid profile found and compared in published literature is also dominated by these PUFAs. As an example, it was found that for canola oil the proportion of  $\alpha$ -linolenic acid relative to linoleic acid was of approximate 2.24:1, which is close to the ideal 2:1 human consumption proportion that has been reported for this crop (Shahidi, 1990, chap. 6).

### 3.2. Mass spectrometry analysis

The full scan spectra in the negative ion mode of all fatty acid analytes were dominated by the  $[M-H]^-$  anions as base peaks and for that reason these were taken as precursor ions in order to conduct further fragmentation studies by tandem mass spectra. Table 3 shows the masses of the base peaks observed for ESI in the negative mode.

Data dependent experiments were programmed on the Xcalibur<sup>®</sup> analytical software with the main objective of conducting MS/MS fragmentation of standards and acquire all the possible characteristic product spectra for subsequent structure elucidation and confirmation of analytes in oil samples. Because many compounds can have the same molecular mass, the intact mass of a compound is not a unique identifier and thus ion transitions, in which a parent mass gives a set of particular fragment peaks, can be monitored to decrease the likelihood of false positives results.

The CID MS/MS spectra of saturated fatty acids were dominated by  $[M-H]^-$  peaks and the fragment ions  $[M-18]^-$  produced by a loss of  $H_2O$  from the carboxyl group (Kerwin et al., 1996; Perret

et al., 2004). As stated in the literature, the NI MS/MS spectra of arachidic ( $C_{20:0}$ ) and behenic ( $C_{22:0}$ ) acids also yielded fragment ions at  $m/z$  183 which are explained by charge remote fragmentations (CRFs), a class of gas-phase decompositions that occur physically remote from the charge site and which are analytically useful in the determination of double bond positions and branching in the aliphatic chain (Kerwin et al., 1996; Perret et al., 2004). The spectra of mono-unsaturated standards was also dominated by  $[M-H]^-$  peaks and the fragment ions  $[M-18]^-$  although more fragments could be seen with respect to that of the saturated standards (Fig. 1). A study by Kerwin et al. (1996) showed that the MS/MS fragmentation of monounsaturated compounds in the NI mode of ESI produced an ion spectra very similar to those of saturated fatty acids; however, by limiting the scanning range of the most abundant ions of the spectrum they could increase the signal intensity and, in this way, determine more fragments that could be rationalised by the CRF process described before.

Examination of the  $\alpha$ -(18:3 $\omega$ 6) and  $\gamma$ -(18:3 $\omega$ 3) linolenic acid isomers showed that the most intense fragments peaks from the precursor  $[M-H]^-$  ions were formed by the loss of  $H_2O$  ( $m/z$  259) and  $CO_2$  ( $m/z$  233), this information agrees with published literature (Kerwin et al., 1996). The CID spectra of these poly-unsaturated fatty acids as well as of some mono-unsaturated fatty acids were also characterised by a series of ion peaks each one separated by 14  $m/z$  units, representing, according to Perret et al., 2004, CRFs cleavages of consecutive carbon bonds in the fatty acid chains and which were also useful to localise double bonds as the series were interrupted by gaps at their locations (Fig. 2).

As mentioned previously, some unknown peaks were also observed specifically with  $m/z$  values of 227, 307, 325, 353, 381 and 395. However, these values are in the range of most fatty acids  $m/z$  values found, so, some hypothesis were made about their identities.

For peaks with  $m/z$  227 and 307, it is believed that both are related to the mass of myristic ( $C_{14:0}$ ) and eicosadienoic ( $C_{20:2}$ ) acids respectively. Retention time data of these two unknown peaks in comparison to those of standards also suggested the presence of these acids in the oil of pea varieties 29579 and Big Pea (8.79 min  $\pm$  0.02 for  $m/z$  277 in 29579 compared to 8.71 min  $\pm$  0.02 to the myristic acid standard; and 12.85 min  $\pm$  0.90 for  $m/z$  307 compared to 13.86 min  $\pm$  0.04 for the eicosadienoic acid standard). Some authors have reported relative low percentage values of myristic and

**Table 3**

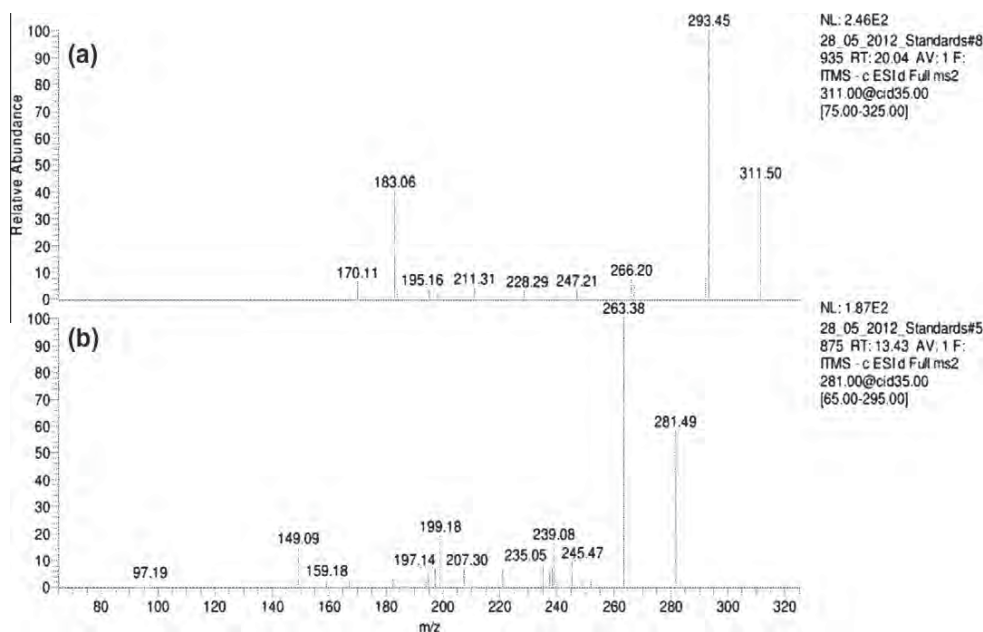
Exact molecular mass,  $[M-H]^-$  ions observed, retention times (RT) and limits of detection (LODs) and quantification (LOQs) for fatty acid standards analysed by RP-LC/ESI-MS.

Compound	Exact masses	$[M-H]^-$ ions detected <sup>a</sup>	RT (min) <sup>a</sup>	LODs (ng) <sup>b</sup>	LOQs (ng) <sup>b</sup>
<i>Saturated fatty acid standards</i>					
Myristic ( $C_{14:0}$ )	228.20	227.20 $\pm$ 0.02	8.71 $\pm$ 0.02	24.06	80.20
Palmitic ( $C_{16:0}$ )	256.24	255.25 $\pm$ 0.03	12.38 $\pm$ 0.04	26.41	88.06
Stearic ( $C_{18:0}$ )	284.27	283.29 $\pm$ 0.02	16.28 $\pm$ 0.04	23.16	77.22
Arachidic ( $C_{20:0}$ )	312.30	311.32 $\pm$ 0.02	19.77 $\pm$ 0.05	5.46	18.20
Behenic ( $C_{22:0}$ )	340.33	339.38 $\pm$ 0.02	23.97 $\pm$ 0.10	28.50	95.01
Lignoceric ( $C_{24:0}$ )	368.36	367.41 $\pm$ 0.02	29.56 $\pm$ 0.05	22.69	75.66
<i>Monounsaturated fatty acid standards</i>					
Palmitoleic ( $C_{16:1}$ )	254.22	253.22 $\pm$ 0.03	9.41 $\pm$ 0.04	22.95	76.53
Oleic ( $C_{18:1}$ )	282.25	281.26 $\pm$ 0.03	13.09 $\pm$ 0.05	19.55	65.17
Gadoleic ( $C_{20:1}$ )	310.28	309.34 $\pm$ 0.03	17.02 $\pm$ 0.06	24.39	81.33
Erucic ( $C_{22:1}$ )	338.31	337.37 $\pm$ 0.04	20.12 $\pm$ 0.07	2.77	9.26
Nervonic ( $C_{24:1}$ )	366.34	365.38 $\pm$ 0.04	24.41 $\pm$ 0.08	9.82	32.75
<i>Polyunsaturated fatty acid standards</i>					
Linoleic ( $C_{18:2}$ )	280.24	279.26 $\pm$ 0.03	10.31 $\pm$ 0.02	15.13	50.45
$\alpha$ -Linolenic ( $C_{18:3}$ )	278.22	277.25 $\pm$ 0.03	8.11 $\pm$ 0.03	17.82	59.42
$\gamma$ Linolenic ( $C_{18:3}$ )	278.22	277.27 $\pm$ 0.02	8.44 $\pm$ 0.02	22.99	76.65
Eicosadienoic ( $C_{20:2}$ )	308.27	307.30 $\pm$ 0.01	13.86 $\pm$ 0.04	23.20	77.36

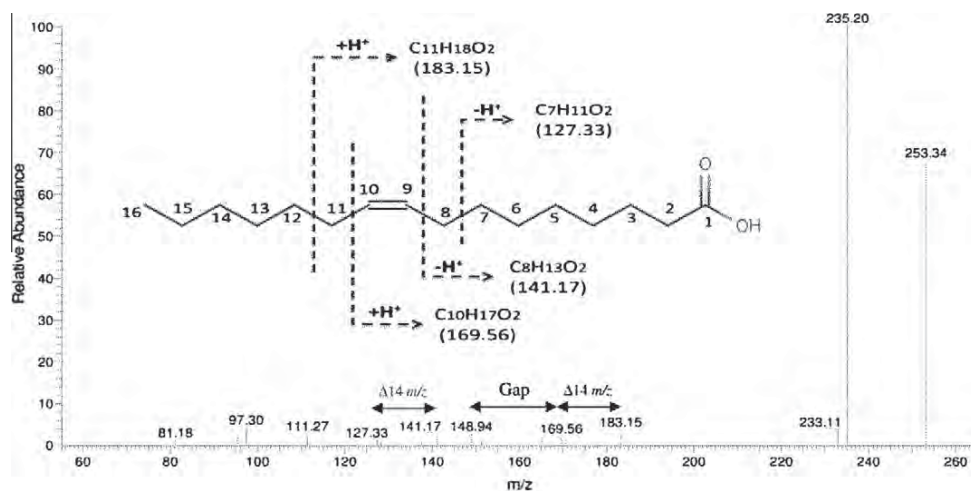
<sup>a</sup> Values presented are the means of five determinations  $\pm$  standard deviation.

<sup>b</sup> Values based on the equation:  $(SD/m) \times 3.3$  or 9. Where SD is the standard deviation of the response (peak areas) and  $m$  is the slope of each calibration curve at levels approximating the LOD or LOQ, respectively.





**Fig. 1.** Negative ion ESI-MS/MS of (a) arachidic ( $C_{20:0}$ ) and (b) oleic ( $C_{18:1}$ ) acids. The parent ions yielding each spectra are the  $[M-H]^-$  peaks detected in full scan experiments and these were also seen here as  $m/z$  311.50 and 281.49. For arachidic acid, peaks at  $m/z$  293.45 and at  $m/z$  183.05 can be explained by the loss of water from the carboxyl moiety and by CRFs processes respectively. For oleic acid, peak at  $m/z$  263.38 can be rationalised by the loss of water from the carboxyl moiety.



**Fig. 2.** NI ESI-MS/MS spectrum of palmitoleic acid ( $C_{16:1}$ ). Peak at  $m/z$  235.20 can be explained by loss of water from the carboxyl moiety and peak  $m/z$  253.34 is the  $[M-H]^-$  ion previously observed. An ion series with an inter-peak spacing of 14  $m/z$  units was also appreciated and can be rationalised by cleavage of consecutive C–C single bonds in the fatty acid chain as shown in the inset structure. The peak at  $m/z$  169.56 was used as an indicator of the presence of a double bond in the molecule.

eicosadienoic acids in different varieties of field pea supporting the present hypothesis (Coxon & Wright, 1985; El-Saied et al., 1981; Yoshida et al., 2007). However, for this study a solid conclusion cannot be given at this time, since the lack of MS/MS spectra for these two peaks did not allow the positive identification of the peaks as real fatty acids or false positives results.

Identity of peaks with  $m/z$  values close to 381 and 395 which were seen in some pea oil samples and in soybean and canola oil are believed to be related to pentacosylic ( $C_{25:0}$ ) and cerotic ( $C_{26:0}$ ) acids respectively. Since no standards of these acids were purchased and the MS/MS spectra of both peaks showed a similar pattern to those of saturated fatty acids, quick searches in the online available databases LIPID MAPS MS Prediction Tool ([\[www.lipidmaps.org/tools/\]\(http://www.lipidmaps.org/tools/\)\) and the Scripps Center for Metabolomics METLIN Database \(<http://metlin.scripps.edu/>\) were also conducted in order to corroborate the assumption. The searches, although not entirely accurate, included in their many results these two fatty acids. Retention times observed for these two peaks in the different samples, eluting almost at 30 min and always after lignoceric acid \( \$C\_{24:0}\$ \) and eluting first  \$m/z\$  381 and then 395 are also keys for the probable identity of these peaks.](http://</a></p>
</div>
<div data-bbox=)

Some authors have reported the presence of very long chain fatty acids (VLFAs) in some species of the family Leguminosae. Sayed, Afifi, and Hassan (1980) reported that *Vicia sativa* subsp. *nigra*, *Vicia calcarata* and *Vicia faba* var. *minor* contain considerable amounts of higher molecular weight saturated fatty acids.



Castellón et al. (2003) also analysed the biochemical composition of the seed of six cowpea cultivars and found low percentages of pentacosanoic and arachidic acids. Other example includes groundnut oil, which according to Dean, Davis, and Sanders (2011) contains cerotic acid into its composition. No information regarding the presence of VLFA in Canola was found.

Finally the peaks with  $m/z$  values of 325 and 353 detected in pea oil samples from seeds of varieties 45760 and Galena were not determined as of lack of standards and MS/MS data.

Nevertheless, with this identification attempt some peaks were still used to localise double bonds, distinguish between positional isomers such as  $\alpha$ - and  $\gamma$ -linolenic acid and make more accurate recognitions of fatty acids than with just a single stage MS experiment.

### 3.3. RP-LC of fatty acids

The elution pattern for each fatty acid was influenced by the carbon chain length and degree of unsaturation with long chain fatty acids having higher RT than short chain fatty acids but this effect was partially offset by the effect of decreasing RT with increasing degree of unsaturation (Chen & Chuang, 2002; Seiwert et al., 2009).

The most difficult fatty acid pairs resolved were the positional isomers of linolenic acid ( $C_{18:3}$ ) with myristic acid ( $C_{14:0}$ ) followed by the behenic ( $C_{22:0}$ ) and nervonic ( $C_{24:1}$ ) acids pair. Retention times of fatty acids may be predicted by semi-empirical means (Chen & Chuang, 2002). It is often believed that the equivalent chain length (ECL) of fatty acids with  $N$  carbon atoms and  $n$  double bonds can be expressed as:

$$ECN = \text{Chain length} - (2)(\text{Number of double bonds})$$

Components with the same ECN will elute closely to each other and be considered as “critical pairs” for the separation (Purdon, 1993). According to Chen and Chuang (2002) each double bond produces a retention time approximating that of two fewer methylene groups; this therefore results in close elution of fatty acids such as  $C_{14:0}$ ,  $C_{16:1}$ ,  $C_{18:2}$  and so on. These “critical pairs” also include  $C_{16:1}$  and  $C_{20:4}$ , cis and trans  $C_{18:1}$ ,  $C_{20:0}$  and  $C_{22:1}$ , and  $C_{22:0}$  and  $C_{24:1}$ . Although there have been several attempts to separate

critical pairs of fatty acids, simultaneous separation of these molecules with different chain lengths and unsaturation in addition to cis/trans isomers has been proven difficult (Purdon, 1993).

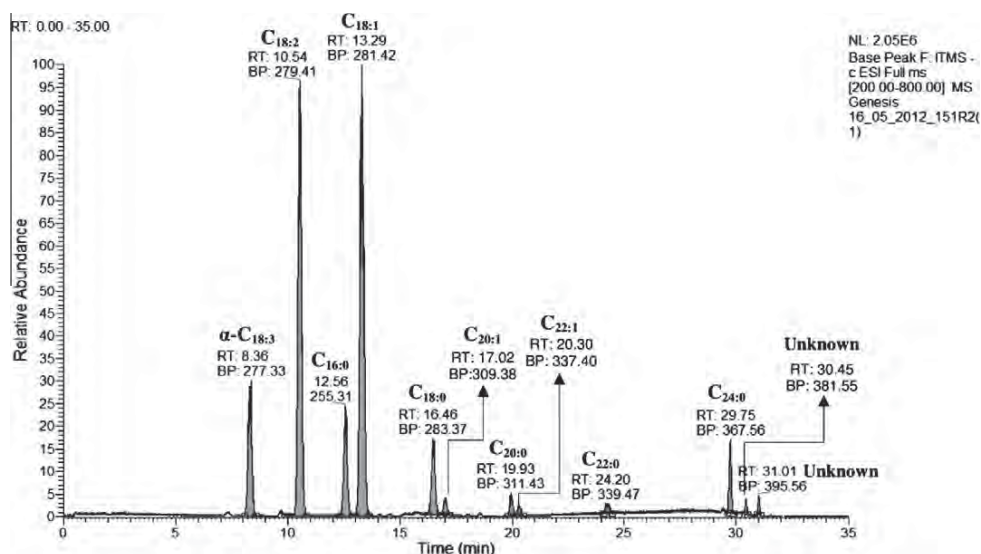
Separation based on double bond position can also be successful in limited cases. For example, oleic acid ( $C_{18:1}$ , 9c) can be separated from petroselinic acid ( $C_{18:1}$ , 6c), but is not usually separated from vaccenic acid ( $C_{18:1}$ , 11c) (Purdon, 1993). Identification of positional isomers of linolenic acid by MS has been reported as unsatisfactory and therefore chromatographic separation is necessary (Laakso, 1997). In the present study the isomers of linolenic acid tested were partially resolved between each other with  $\alpha$ -linolenic acid molecules eluting first then the corresponding  $\gamma$ -linolenic acid.

The HPLC elution strategy employed in this work was based on the studies by Li et al. (2001) and Chen and Chuang (2002). In Li et al. (2001) a gradient from 100% mobile phase A (25% acetonitrile + 75% water + 0.12% v/v acetic acid) to 100% mobile phase B (100% acetonitrile + 0.12% v/v acetic acid) linearly over 80 min was acceptable to separate the critical pairs up to  $C_{18:3}$ , however when trying to separate longer carbon chain fatty acids the elution failed (data not shown). Chen and Chuang (2002) investigated the effect of temperature and organic solvent composition on the RP separation of coumarin derivatised fatty acids according to their carbon number (from  $C_{14}$  to  $C_{22}$ ), the degree of unsaturation and their trans/cis configuration and found that methanol as organic modifier is able to separate long chain fatty acid pairs and structural isomers by altering the ECL values of the molecules. For that reason, the elution program also included methanol as a third mobile phase.

The capability of the proposed method was examined with the pea, soybean and canola samples previously prepared. Fig. 3 is a representative example of the separation of fatty acids achieved in pea oil.

Individual fatty acids were identified based on the MS/MS spectra patterns and on retention times. The isomer  $\alpha$ - $C_{18:3}$  was detected and is in general agreement with compositions of pea, soybean and canola oil that have been investigated using standard GC-FID or GC-MS techniques (Hammond et al., 2005; Shahidi, 1990, chap. 6; Yoshida et al., 2007).

Although retention times of the standards were different than those in the biological samples the same elution pattern was observed. This shift in retention can be due to changes in the pH of



**Fig. 3.** Separation of fatty acids from the oil of field pea 29595. In all screened pea samples a similar fatty acid profile was appreciated. No relevant contaminant peaks were seen. Scanned mass range was from 200.00 to 800.00. Identifications of individual fatty acids were based on MS/MS spectrum profile and retention times.

the mobile phases or due to the number of compounds present in the standard solution in opposition to the samples as peak spacing can change if more compounds of similar structure are present.

#### 4. Linearity of the response detector and limits of detection

The response of the system was established by measuring peak areas for fatty acid standards over a 5-fold range of concentration. This response was linearly dependent on the amounts of analyte up to 100 ng, with each standard calibration curve showing  $R^2$  values of more than 0.90.

Limits of detection (LODs) and quantification (LOQs) were calculated based on the standard deviation of the response (peak areas) and the slope values of each standard calibration curve and by multiplying the quotient of both by 3.3 or 9 for LODs and LOQs respectively. These data are presented in Table 3.

#### 5. Conclusions

A simple and rapid methodology by LC/ESI-MS/MS for the characterisation of fatty acids from pea oil samples was developed. The procedure was also tested with soybean and canola oils showing the possible application for oilseed investigation of fatty acids. Sample preparation and manipulation was minimal and extracts were clean enough for analysis. Compared to methods such as GC-FID or LC-UV/Vis, chemical derivatisation was not necessary thus decreasing the likelihood of unwanted effects like racemisation and/or oxidation. Generally in raw pea the following fatty acid composition was obtained: C18:2 > C18:1 > C16:0 > C18:3 > C18:0, which is consistent with literature. This fatty acid profile suggests that the species of field pea investigated may have the potential to be used as raw materials to develop a commercially accepted oilseed crop as oil composition of field pea perfectly fits for the food industry.

#### Acknowledgments

This work is being supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) (project no. 120497570), the Consortium de Recherche et Innovations en Bioprocédés Industriels au Québec (CRIBIQ), Lefsrud Seed and Processors Ltd and Agrocentre Belcan.

Special thanks to Dr. Eric Huang, McGill University, for helping with the technical and software handling of the LC-MS equipment used.

Manuel Villalobos would like to thank the Mexican National Council on Science and Technology (CONACYT) and the Fund for the Development of Human Resources (FIDERH) for financing his studies in Canada.

#### References

- Bağcı, E., & Şahin, A. (2004). Fatty acid patterns of the seed oils of some *Lathyrus* species L. (Papilionideae) from Turkey, a chemotaxonomic approach. *Pakistan Journal of Botany*, 36, 403–413.
- Balogun, A., & Fetuga, B. (1985). Fatty acid composition of seed oils of some members of the Leguminosae family. *Food chemistry*, 17(3), 175–182.
- Byrdwell, W. C. (2005). *Modern methods for lipid analysis by liquid chromatography/mass spectrometry and related techniques*. AOCS Press.
- Cahoon, E. B. (2003). Genetic enhancement of soybean oil for industrial uses: Prospects and challenges. *Journal of Agrobiotechnology Management & Economics*, 6(1–2), 11–13.
- Castellón, R. E. R., Araújo, F. M. M. C., Ramos, M. V., Andrade Neto, M., Freire-Filho, F. R., Grangeiro, T. B., et al. (2003). Biochemical composition and characterisation of lipid fraction of six cowpea cultivars. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 7(1), 149–153.
- Chen, S. H., & Chuang, Y. J. (2002). Analysis of fatty acids by column liquid chromatography. *Analytica Chimica Acta*, 465(1–2), 145–155.
- Coxon, D. T., & Wright, D. J. (1985). Analysis of pea lipid content by gas chromatographic and microgravimetric methods. Genotype variation in lipid content and fatty acid composition. *Journal of the Science of Food and Agriculture*, 36(9), 847–856.
- Dean, L. L., Davis, J. P., & Sanders, T. H. (2011). Groundnut (peanut) oil. In F. D. Gunstone (Ed.), *Vegetable oils in food technology: Composition, properties and uses* (pp. 225–242). Oxford: Blackwell Publishing Ltd.
- El-Saied, H. M., Amer, M., & Gabran, A. (1981). Unsaponifiable matter and fatty acid composition of pea oil. *Zeitschrift für Ernährungswissenschaft*, 20(2), 132–138.
- Hammond, E. G., Johnson, L. A., Su, C., Wang, T., & White, P. J. (2005). *Soybean oil*. Bailey's Industrial Oil and Fat Products.
- Hilditch, T. P., & Williams, P. N. (1964). *The chemical constituents of natural fats* (4th ed.). London: Chapman and Hall.
- Holman, R. T. (1981). Essential fatty acids in nutrition and disease. *Chemistry & Industry*, 21, 704–709.
- Kerwin, J. L., Wiens, A. M., & Ericsson, L. H. (1996). Identification of fatty acids by electrospray mass spectrometry and tandem mass spectrometry. *Journal of mass spectrometry*, 31(2), 184–192.
- Kurata, S., Yamaguchi, K., & Nagai, M. (2005). Rapid discrimination of fatty acid composition in fats and oils by electrospray ionization mass spectrometry. *Analytical sciences*, 21(12), 1457–1465.
- Laakso, P. (1997). Characterisation of  $\alpha$ - and  $\gamma$ -linolenic acid oils by reversed-phase high-performance liquid chromatography-atmospheric pressure chemical ionization mass spectrometry. *Journal of the American Oil Chemists' Society*, 74(10), 1291–1300.
- Li, Z., Gu, T., Kelder, B., & Kopchick, J. (2001). Analysis of fatty acids in mouse cells using reversed-phase high-performance liquid chromatography. *Chromatographia*, 54(7–8), 463–467.
- Lima, E., & Abdalla, D. (2002). High-performance liquid chromatography of fatty acids in biological samples. *Analytica Chimica Acta*, 465(1), 81–91.
- Murcia, M., & Rincon, F. (1992). Size as source of variance in lipid composition of pea. *Food chemistry*, 44(1), 29–35.
- Nikolopoulou, D., Grigorakis, K., Stasini, M., Alexis, M., & Iliadis, K. (2007). Differences in chemical composition of field pea (*Pisum sativum*) cultivars: Effects of cultivation area and year. *Food chemistry*, 103(3), 847–852.
- Nishiyama-Naruke, A., Souza, J., Carneiros, M., & Curí, R. (1998). HPLC determination of underivatized fatty acids saponified at 37 °C analysis of fatty acids in oils and tissues. *Analytical letters*, 31(14), 2565–2576.
- Perret, D., Gentili, A., Marchese, S., Sergi, M., & Caporossi, L. (2004). Determination of free fatty acids in chocolate by liquid chromatography with tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, 18(17), 1989–1994.
- Pulse Canada. Canadian pulse industry: Situation and outlook. (2012). URL: <http://www.agr.gc.ca/pol/maddam/index\_e.php?s1=pubs&s2=spec&s3=php&page=spec\_2012-06-14> Accessed 30.07.12.
- Purdon, M. (1993). Application of HPLC to lipid separation and analysis: Sample preparation. In E. G. Perkins (Ed.), *Analyses of fats, oils and derivatives* (pp. 145). AOCS.
- Ratnayake, W., Hoover, R., Shahidi, F., Perera, C., & Jane, J. (2001). Composition, molecular structure, and physicochemical properties of starches from four field pea (*Pisum sativum* L.) cultivars. *Food chemistry*, 74(2), 189–202.
- Ryan, E., Galvin, K., O'Connor, T., Maguire, A., & O'Brien, N. (2007). Phytosterol, squalene, tocopherol content and fatty acid profile of selected seeds, grains, and legumes. *Plant Foods for Human Nutrition (Formerly Qualitas Plantarum)*, 62(3), 85–91.
- Sayed, D., Afifi, M. S., & Hassan, M. A. (1980). A study of lipid content of the leaves, stems and seeds of *Vicia sativa* L., *Vicia calcarata* Desf. and *Vicia faba* L. (var. minor) growing in Egypt. *Egyptian Journal of Pharmaceutical Sciences*, 21(12), 43–52.
- Seiwert, B., Givalisco, P., & Willmitzer, L. (2009). Advanced mass spectrometry methods for analysis of lipids from photosynthetic organisms. In H. Wada & N. Murata (Eds.), *Lipids in photosynthesis* (pp. 445–461). Springer.
- Shahidi, F. (1990). *Canola and rapeseed: Production, chemistry, nutrition, and processing technology* (1st ed.). New York: Van Nostrand Reinhold.
- Singh, N., Kaur, N., Rana, J. C., & Sharma, S. K. (2010). Diversity in seed and flour properties in field pea (*Pisum sativum*) germplasm. *Food chemistry*, 122(3), 518–525.
- Thiessen, D. L. (2004). Optimisation of feed peas, canola and flaxseed for aqua feeds: The Canadian prairie perspective. *Avances en Nutricion Acuicola VII. Memorias del VII Simposium Internacional de Nutricion Acuicola* (pp. 16–19).
- Wang, N., & Daun, J. K. (2004). *The Chemical composition and nutritive value of Canadian pulses* (pp. 19–29). Canadian Grain Commission (CGC).
- Welch, R. W., & Wynne Griffiths, D. (1984). Variation in the oil content and fatty acid composition of field beans (*Vicia faba*) and peas (*Pisum spp.*). *Journal of the Science of Food and Agriculture*, 35(12), 1282–1289.
- Yoshida, H., Tomiyama, Y., Tanaka, M., & Mizushima, Y. (2007). Characteristic profiles of lipid classes, fatty acids and triacylglycerol molecular species of peas (*Pisum sativum* L.). *European Journal of Lipid Science and Technology*, 109(6), 600–607.



## Effect of nonionic surfactant Brij 35 on the fate and transport of oxytetracycline antibiotic in soil

Eman M. ElSayed\*, Shiv O. Prasher\*\*, Ramanbhai M. Patel

Department of Bioresource Engineering, McGill University, Ste Anne de Bellevue, H9X 3V9 Quebec, Canada

### ARTICLE INFO

#### Article history:

Received 6 July 2012

Received in revised form

14 November 2012

Accepted 23 November 2012

Available online 5 January 2013

#### Keywords:

Nonionic surfactants

Brij 35

Oxytetracycline

Antibiotic

Wastewater

Fate

Lysimeter

### ABSTRACT

In many parts of the world, river water is used for irrigation. Treated, partially treated, and even untreated water from wastewater treatment plants is discharged directly into rivers, thereby degrading the quality of the water. Consequently, irrigation water may contain surfactants which may affect the fate and transport of chemicals such as pesticides and antibiotics in agricultural soils. A field lysimeter study was undertaken to investigate the effect of the nonionic surfactant, Brij 35, on the fate and transport of an antibiotic, Oxytetracycline, commonly used in cattle farms. Nine PVC lysimeters, 1.0 m long  $\times$  0.45 m diameter, were packed with a sandy soil to a bulk density of 1.35 Mg m<sup>-3</sup>. Cattle manure, containing Oxytetracycline, was applied at the surface of the lysimeters at the recommended rate of 10 t/ha. Each of three aqueous Brij 35 solutions, 0, 0.5 and 5 g L<sup>-1</sup> (i.e., 'good,' 'poor' and 'very poor' quality irrigation water) were each applied to the lysimeters in triplicate. Over a 90 day period, soil and leachate samples were collected and analyzed. Batch experiment results showed that the presence of the nonionic surfactant Brij 35 significantly reduced the sorption coefficient of OTC from 23.55 mL g<sup>-1</sup> in the aqueous medium to 19.49, 12.49 and 14.53 in the presence of Brij 35 at concentrations of 0.25, 2.5 and 5 g L<sup>-1</sup>, respectively. Lysimeter results indicated the significant downward movement of OTC at depths of 60 cm into soil profile and leachate in the presence of surfactant. Thus, the reuse of wastewater containing surfactants might enhance the mobility of contaminants and increase ground water pollution.

© 2012 Elsevier Ltd. All rights reserved.

### 1. Introduction

Given the scarcity of water in many parts of the world, the reuse of wastewater has become an important alternative option for irrigation. In developing countries, sometimes wastewater is used to irrigate crops and vegetables; this wastewater may be untreated, or partially treated, or both. Chatterjee (2008) indicated that 85% of 53 cities in developing countries discharge untreated or primary treated wastewater to irrigate land in order to keep pace with the ever increasing demand for food resulting from a growing population. As a consequence of this practice, there may be an increase in bacterial infections, as well as roundworm and diarrheal diseases. A recent survey indicates that 46 countries representing 75% of the world's irrigated land make use of polluted water for irrigation (Drechsel et al., 2010). Globally, Hamilton et al. (2007) estimated that approximately 20 million hectares of land are

irrigated with wastewater, and this area is expected to markedly increase in the next few years.

Wastewater is a mixture of inorganic and organic substances originating in the discharge of municipal and industrial effluents. Many studies reveal that surfactants have been detected as the highest concentrations found among the organic chemicals in wastewater and sludge (Brunner et al., 1988; Field et al., 1992; Wild et al., 1990). Non-ionic surfactants concentrations ranging from 10<sup>2</sup> to 10<sup>3</sup> mg L<sup>-1</sup> have been detected in wastewater (Narkis and Ben-David, 1985). Such nonionic surfactants as alcohol ethoxylates (AEOs) and alkylphenol ethoxylate (APEO) are being used extensively to remove skin grease in the leather industry. Nonionic surfactants were the major pollutants in untreated wastewater samples from the textile industries (Castillo et al. 1999). Although the treatment of effluent in wastewater treatment plants (WWTPs) eliminates the surfactants which undergo degradation to metabolites such as nonylphenol ethoxylates (NPEOs), Loyo-Rosales et al. (2007) found that WWTPs do not remove all of the metabolites. Therefore, many household products contribute to an abundance of these contaminants in urban sewers (Conn et al., 2006; Loyo-Rosales et al., 2007; Stuart et al., 2012).

Basically, non-ionic surfactants do not ionize in water because they have a non-dissociable hydrophilic group (e.g., alcohol, phenol,

\* Corresponding author. Tel.: +1 514 463 7799.

\*\* Corresponding author. Tel.: +1 514 398 5557.

E-mail addresses: [Eman.elsayed@mail.mcgill.ca](mailto:Eman.elsayed@mail.mcgill.ca), [emisamir55@yahoo.ca](mailto:emisamir55@yahoo.ca) (E.M. ElSayed), [Shiv.Prasher@mcgill.ca](mailto:Shiv.Prasher@mcgill.ca) (S.O. Prasher).

ether, ester, or amide) and are considered to be amphiphilic compounds. Generally, the presence of nonionic surfactants in irrigation water leads to two major concerns. First, nonionic surfactant byproducts such as NPEO and nonylphenol (NP) are estrogen disrupting compounds. This may result in changes in the sex ratio, a decrease in fish population and malformations have been observed in animals (Esperanza et al., 2004). Additionally, Barber et al. (2009) indicate that nonionic surfactant degradation products NP were detected in wells in the US; in the United States, potable water contamination is of great concern due to the estrogenic effect of these compounds with the continuous disposal of wastewater treatment plant effluent. Therefore, alcohol ethoxylates (AEOs) are now used as an alternative to NPEOs, as they are safer in terms of their rapid degradation to less toxic, less persistent and less estrogenic metabolites. Secondly, nonionic surfactants in irrigation water may have an impact on the mobility of chemicals in soils and result in the contamination of ground water (Huggenberger et al., 1973; Kan and Tomson, 1990; Aronstein et al., 1991; Rodriguez-Cruz et al., 2004). For example, nonionic surfactants at low concentrations ( $0.04 \text{ g kg}^{-1}$ ) decreased the mobility of the herbicide metolachlor in soil; however, at high concentrations ( $5\text{--}50 \text{ g kg}^{-1}$ ) the mobility increased (Sanchez-Camazano et al., 1995). Furthermore, Katagi (2008) reported that the mobility of pesticides is increased in leached water when there are high concentrations of surfactants. Generally, nonionic surfactants increased the downward movement of pesticides in soil whereas cationic surfactants markedly decreased the mobility of pesticides in the soil. Nilufer (2005) stated that at low concentrations ( $12 \text{ mg L}^{-1}$ ) of LAS and NP, there is no effect on the mobility of such herbicides as atrazin, metolachlor and metribuzin. However, Tao et al. (2006) indicated that at low concentrations  $0\text{--}20 \text{ mg L}^{-1}$ , surfactants have the potential to decrease sorption and mobilize herbicides such as atrazin.

Agricultural effluents are known as a main, non-point source for pollution with pharmaceutically active compounds (PhACs) and their metabolites. Pharmaceutical antibiotics are widely used in agriculture not only to control disease but also as a growth promoter in the livestock industry (Thiele Bruhn, 2003). Application of manure from such industries on agricultural land is of great concern (Stuart et al., 2012). Recent studies revealed that manure may contain as much as 60–90% of the antibiotics excreted in animal feces in the active form due to its poor metabolism. As a consequence, the inactive antibiotic conjugates could be bioactive in manure by cleaving the acetyl group, resulting in the release of the original active form. They can be more persistent in the environment due to their high polarity. Martinez-Carballo et al. (2007) reported that oxytetracycline (OTC) has been detected in manure in concentrations as high as  $29 \mu\text{g g}^{-1}$ . Similarly, Uslu and Balcioglu (2009) found tetracyclines in manure samples around a concentration of  $46 \mu\text{g g}^{-1}$ . Antibiotics may enter the environment via different pathways such as agricultural soil, ground and surface water (Halling-Sorensen et al., 1998). OTC is considered the preferred tetracycline antibiotic in fish farms and as an animal feed additive; given that fish farm effluent lacks additional treatments to reduce OTC residue, OTC residues could be introduced into water bodies easily as a result of the discharge from fish farms (Himmelsbach and Buchberger, 2005). The sorption coefficient ( $K_d$ ) values of OTC are found to be between 416 and  $1026 \text{ L/kg}$  in sandy and sandy loam soil, respectively (Rabolle and Spliid, 2000). Sorption of OTC to clay loam and loam sand soil was tested as a function of pH and ionic strength and the results show that as pH and ionic strength increase, the sorption decreases (Terlaak et al., 2006). Similarly, as dissolved organic matter (DOM) increases, OTC sorption decreases leading to increasing mobility; this suggests that DOM is an important factor affecting OTC leaching in soil (Kulshrestha et al., 2004).

Terlaak et al. (2006) stated that manure contains high amounts of ammonium nitrate which is transformed to nitrate by releasing protons, and as a result, it decreases soil pH. Therefore, antibiotics adsorption potential onto soil varies considerably after manure application and the chemical interaction of manure with the soil. Additionally, Wang and Yates (2008) estimated the half-life of OTC to be 33 days in manure-amended soil compared to 8.1 and 56 days in manure and non-amended soil, respectively; suggesting that OTC persistence in soil may increase after manure application to soil. Furthermore, the half-life of OTC in soil under aerobic conditions ranged between 26 and 56 days in non-sterile soil, whereas the half-life increased to 99–120 days in sterile soils. This confirms the fact that microorganisms can degrade OTC. Nevertheless, abiotic degradation is important in the dissipation process in soil (Yang et al., 2009). Blackwell et al. (2009) investigated the fate of OTC in a sandy loam soil column under extreme rainfall conditions and found that OTC was detected in soil samples only from the 0 to 5 cm top of the lysimeter profile; OTC was not found in the leachate due to the high sorption coefficient and its persistence in soil. Similar results were obtained by Rabolle and Spliid (2000) and Blackwell et al. (2007). However, Hamscher et al. (2000) analyzed the leachate samples collected from agricultural land and reported that chlortetracycline, oxytetracycline, tetracycline and tylosin were detected at concentrations ranging from 0.1 to  $0.3 \mu\text{g L}^{-1}$ . This could pose a risk of ground water contamination and bacterial resistance. On the other hand, Boxall et al. (2006) reported that OTC significantly reduced plant growth and similar effects have been documented with other groups of pharmaceuticals as well. Antibiotics are also considered as emerging contaminants; they are of great concern because of their potential to enhance the rise of antibiotic resistant genes. A prolonged exposure to low doses of antibiotics may lead to an increase in drug resistance among the soil microbial community. Moreover, Rhodes et al. (2000) found evidence of the likelihood of the transfer of antibacterial resistance to humans.

From the literature, it is evident that when manure containing oxytetracycline is applied to agricultural land, oxytetracycline may enter the water system and then surfactants may affect its mobility. To the best of the authors' knowledge, no study has been done to address the fate of oxytetracycline when it coexists with the nonionic surfactants in soil. Therefore, this study was conducted to investigate the mobility of oxytetracycline in soil (with low organic matter content) and water in the presence of nonionic surfactant Brij 35 in the irrigation water.

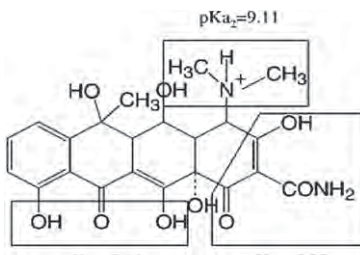
## 2. Material and methods

### 2.1. Study compounds

Oxytetracycline (OTC) hydrochloride 95% HPLC grade from Sigma–Aldrich, solubility  $1 \text{ g L}^{-1}$  (Table 1) and Oxyvet as a commercial grade of OTC used in cattle farms was purchased from CMFO, Quebec. Demeclocycline (DMCTC) was used as an internal standard (to better estimate the loss of OTC during extraction procedure from soil and leachate samples) and it was purchased from Sigma–Aldrich. A pure standard (Pestanol<sup>®</sup>) and formulation (Sencor 75) of metribuzin were purchased from Sigma–Aldrich. The non-ionic surfactant Brij 35 [Polyoxyethylene lauryl ether ( $\text{C}_2\text{H}_4\text{O}$ )<sub>23</sub>– $\text{C}_{12}\text{H}_{25}\text{OH}$ ] with critical micellar concentration (CMC) =  $74 \text{ mg L}^{-1}$  and  $\text{Log } P$  is approximately 3.13 (HERA, 2009). HPLC-grade methanol and acetonitrile were obtained from Sigma–Aldrich. Mobile phase chemicals were purchased from Fisher Scientific. Double-deionized water (Milli-Q, Millipore, Molsheim, France) was used in the preparation of standard solutions and mobile phase solutions.



**Table 1**  
Physical and chemical characteristics of oxytetracycline.

Physical–chemical properties	Chemical structure
Solubility: 1 g L <sup>-1</sup> MW: 460.44 Log <i>K</i> <sub>ow</sub> : -1.22 <sup>a</sup> GUS index: -0.91 <sup>b</sup> p <i>K</i> <sub>a1</sub> : 3.27 <sup>c</sup> p <i>K</i> <sub>a2</sub> : 7.32 p <i>K</i> <sub>a3</sub> : 9.11	

<sup>a</sup> Loke et al. (2002).

<sup>b</sup> PPDB (2009).

<sup>c</sup> Sassman and Lee (2005).

## 2.2. Soil characteristics

Belonging to the St. Amble complex, the sandy soil used was obtained from a field in the Macdonald Campus of McGill University, Ste-Anne-De-Bellevue, Qc. Physical and chemical properties of the soil are given in Table 2.

### 2.2.1. Batch adsorption experiment

Adsorption of OTC onto soil in the presence of the nonionic surfactant Brij 35 was investigated through a soil equilibrium technique. Triplicate soil samples (1 g) were equilibrated with an aqueous solution (20 mL) of OTC at concentrations ranged from 5 to 50 µg mL<sup>-1</sup> in factorial combination with Brij 35 concentrations of 0, 0.02, 0.25, 0.5 and 5 g L<sup>-1</sup>. All solutions contained 0.01 M CaCl<sub>2</sub> to mimic the ionic strength of ground water (Wilde et al., 2008). The soil slurry was placed in a 50 mL centrifuge tube, and placed on a shaker, for 48 h. The slurry was centrifuged at 2683 g for 20 min. Extraction and analysis of OTC are then proceeded. The proportion of OTC adsorbed was calculated by taking the difference between the amount initially present in solution and the amount remaining in solution after equilibrium with the soil.

### 2.2.2. Analytical procedure

The soil samples and leachate were analyzed by HPLC with diode array detection DAD using eclipse plus C18 column (4.6 × 150 mm) from Agilent, USA. The mobile phase was 15% acetonitrile and 85% of water containing phosphate buffer at pH 2.5. The flow rate was 1 mL min<sup>-1</sup> and detection wavelength was at 355 nm for OTC and 254 nm for DMCTC with detection limit of 0.1 ng g<sup>-1</sup>.

OTC sorption data were fitted to the Freundlich adsorption isotherm equation:

$$q = K_f C_{aq}^n \quad (1)$$

where *q* is the amount of OTC sorbed (µg g<sup>-1</sup>), *C*<sub>aq</sub> is OTC concentration after equilibrium in µg mL<sup>-1</sup>, and *K*<sub>f</sub>, *n* are the Freundlich affinity and nonlinearity coefficients, respectively (mL g<sup>-1</sup>).

**Table 2**  
Physical and chemical characteristics of soil.

Soil type	Sand (%)	Silt (%)	pH	Bulk density (Mg m <sup>-3</sup> )	Organic matter (%)	CEC (cmol kg <sup>-1</sup> )	Hydraulic conductivity (cm d <sup>-1</sup> )	The zero point of charge (zpc)
Sandy	92.2	4.3	5.5	1.350	2.97	4.9	1.67 (S.D. = 0.45)	3.4

Organic carbon normalized coefficient values for oxytetracycline in the absence of surfactant and in the presence of surfactant were obtained as follows:

$$K_{oc} = \frac{K_f}{f_{oc}} \quad (2)$$

where *f*<sub>oc</sub> is the organic carbon fraction of the soil (0.0297 g g<sup>-1</sup> for the soil used), and *K*<sub>oc</sub> is the partition coefficient of the contaminant in the organic fraction of the soil (mL g<sup>-1</sup>).

## 2.3. Field experimental set up

The field experiment investigating oxytetracycline transport in a sandy agricultural soil was conducted in nine outdoor PVC lysimeters set up at the Macdonald Campus of McGill University, Ste-Anne-De-Bellevue, Quebec. The lysimeters (0.45 m O.D. × 1 m long), were sealed at the bottom to 0.6 m × 0.6 m PVC sheets. Each lysimeter was packed in layers with sandy soil and adjusted to a bulk density of 1.35 Mg m<sup>-3</sup>. A 0.05 m diameter drainage pipe was installed at the bottom of each lysimeter. Four, 10 mm diameter, soil sampling holes were made in each lysimeter at depths of 0.1, 0.3 and 0.6 m the soil surface (Fig. 1). The lysimeters were sheltered to prevent the entry of natural precipitation.

### 2.3.1. Application of tested compounds

Antibiotic free manure was collected from a cattle farm in Ham-Sud, Quebec. The manure samples were mixed before adding the antibiotic. At the end of June 2009, before the start of the experiment, all of the lysimeters were irrigated, bringing them to saturation. The drainage pipe at the bottom of each lysimeter was left open throughout the study. The next day (Day 0), with the lysimeter's soil at field capacity, homogenized manure (10 Mg ha<sup>-1</sup>) was applied to the surface of all of the lysimeters and manually mixed into the top 30 mm of soil (Chettri and Thapa, 2004) after being spiked with oxytetracycline in the form of Oxyvet according to the findings of Arikian et al. (2007). (The spiked concentrations were calculated after adjusting the amount of OTC extracted from manure in Arikian's study (115 µg g<sup>-1</sup>) to the area of the lysimeter and the amount of manure applied.) Metribuzin was then sprayed on each lysimeter's soil surface (July 1, 2009) at the locally recommended rate for potatoes (1 kg a.i. ha<sup>-1</sup>) as a common practice. Irrigation water (70 mm), containing one of three different concentrations of Brij 35 (0, 0.5 or 5.0 g L<sup>-1</sup>) was applied at day 0, day 21, day 42, day 63, according to the recommended irrigation practice for potatoes. Each treatment was replicated three times. While these surfactant concentrations may appear to be high concerning the environmental concentrations, Abu-Zreig et al. (1999) have studied the effect of surfactants on atrazine movement in soils using concentrations up to 10 g L<sup>-1</sup>.

### 2.3.2. Soil and leachate sampling

Soil samples were collected at the surface as well as at depths of 0.1, 0.3 and 0.6 m, through 4 sampling ports in the side of the lysimeters. Soil samples were collected on 8 occasions: 0, 1, 5, 11, 22, 43, 60 and 90 days after the application of manure. For each surfactant concentration, three replicate soil samples (3 lysimeters) were taken at each of the four depths, and samples from a common



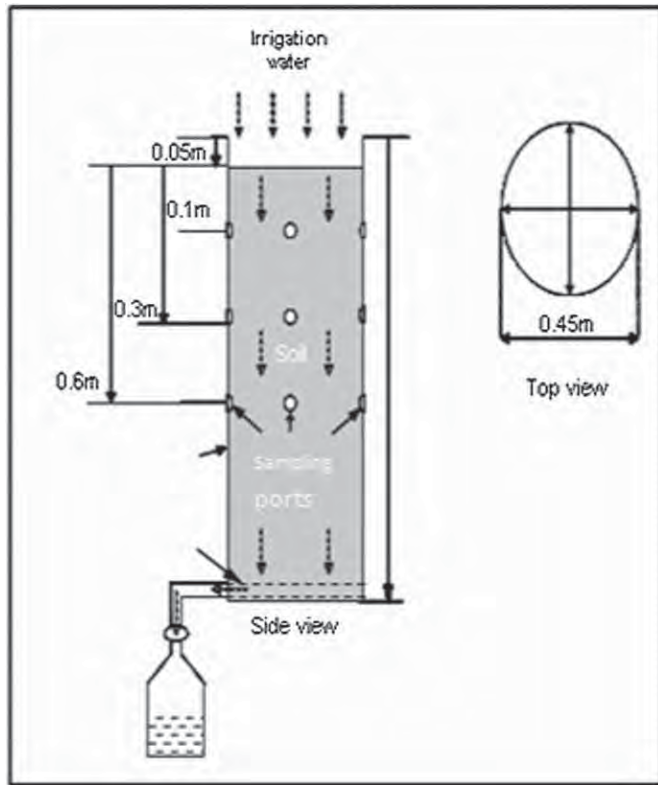


Fig. 1. Schematic design of the lysimeter.

treatment  $\times$  depth were combined in a composite sample ( $\sim 15$  g). Subsamples of about 5 g were taken from each composite sample, and their moisture content was determined. The remaining portions of the samples were stored in sealed bags in the freezer at  $-24$  °C until extraction. Leachate samples (1 L) were collected as a subsample of a total of approximately 11 L of leachate collected at the outlet at the bottom of each lysimeter following each of the four irrigation events. The subsamples were then transported to the lab and immediately extracted to prevent any chance of OTC degradation.

### 2.3.3. Mass balance calculations

At each sampling date, the total antibiotics mass recovered from each lysimeter was calculated as the sum of the antibiotic recovered in soil samples across all depths of the soil profile. Adding the antibiotic mass in the leachate to that in the soil would total the initial antibiotic mass applied, minus any losses (unrecovered oxytetracycline) due to degradation or volatilization (Eq. (1)).

$$\begin{aligned} \text{OTC}_{\text{init}} = & 1590.4\rho \left[ C_{(0-5)}\theta_{(0-5)}h_{(0-5)} + C_{(5-15)}\theta_{(5-15)}h_{(5-15)} \right. \\ & \left. + C_{(15-45)}\theta_{(15-45)}h_{(15-45)} + C_{(45-70)}\theta_{(45-70)}h_{(45-70)} \right] \\ & + [C_{\text{leach}}V_{\text{leach}}] + \text{OTC}_{\text{lost}} \end{aligned} \quad (3)$$

where  $\text{OTC}_{\text{init}}$ , total oxytetracycline (mg) initially applied;  $\text{OTC}_{\text{lost}}$ , oxytetracycline lost through degradation, volatilization, etc. (mg);  $\rho$ , soil bulk density ( $\text{g cm}^{-3}$ );  $C_{(x-y)}\theta_{(x-y)}h_{(x-y)}$ , oxytetracycline concentration and soil moisture content, respectively, in soil layer  $h$  ranging from depth  $x$  to depth  $y$  ( $\text{mg g}^{-1}$ );  $C_{\text{leach}}$ , oxytetracycline concentration in leachate ( $\text{mg L}^{-1}$ );  $V_{\text{leach}}$ , volume of leachate; 1590.4, soil layer surface area,  $\pi(D/2)^2$ , where  $D = 45$  cm.

In order to calculate the mass of oxytetracycline recovered in the soil layer samples, the following equation was used for this purpose

$$\text{Mass of OTC} = [C(m)] * \rho * a * h \quad (4)$$

where  $C$ , laboratory reported analytical oxytetracycline concentration in soil samples ( $\text{mg g}^{-1}$ ),  $m$ , moisture content of soil samples (mass water/mass dry soil  $\text{mg g}^{-1}$ ),  $\rho$ , soil bulk density ( $\text{g cm}^{-3}$ ),  $a$ , area of lysimeter ( $\text{m}^2$ ) and  $h$ , sampling depth of soil layers (cm).

## 2.4. Extraction from leachate and soil samples

### 2.4.1. Leachate samples extraction

Leachate was extracted right after it was transported from the lysimeters. Each 1 L subsample was filtered through 90 mm filter paper, followed by filtration through a 45 mm filter (Advantec, Japan), to remove coarse and fine suspended matter, prior to extraction. HLB cartridges from Oasis Co. Ltd, NY were used in the solid phase extraction (SPE). Prior to extraction, 2 g of EDTA were added to the sample followed by the addition of 100  $\mu\text{L}$  of DMCTC. Preconditioning of each cartridge began by passing 5 mL of water through each cartridge twice, then 5 mL of 50:50 water:methanol (v/v), followed by 100% methanol, all at a flow rate of  $10 \text{ mL min}^{-1}$ . The sample filtrate was then passed through the cartridges and then washed twice with 5 mL of water, to elute OTC and DMCTC. Collected in a test tube, the 10 mL of eluate was evaporated under a  $\text{N}_2$  stream, residue redissolved in 1 mL of acetonitrile, then passed through 0.22  $\mu\text{m}$  syringe driven filter (Millex-GV, Japan), and the filtered solution transferred to a vial for HPLC analysis.

### 2.4.2. Soil samples extraction

One gram of soil sample was weighed into av. 50 mL centrifuge tube and 5 mL of extraction buffer (0.1 M Mcllvaine buffer ( $\text{Na}_2\text{HPO}_4$  and citric acid at pH 7)/0.1 M EDTA/methanol 25:25:50 v/v) then DMCTC was added. The tube was vortexed for 30 s and placed in an ultrasonic bath for 15 min. It was then centrifuged at 2683 g for 20 min. Supernatant was decanted and the extraction procedure was repeated twice. The combined supernatants were passed through a 45 mm filter paper, and then evaporated under a  $\text{N}_2$  stream. Samples were redissolved in 1 mL of acetonitrile and passed through a 0.22  $\mu\text{m}$  syringe driven filter prior to HPLC analysis.

## 2.5. Data analysis

A statistical model using repeated measures over time and depth was employed to determine if the oxytetracycline concentrations differed between treatments, over time and with varying depths. Data was analyzed with PROC MIXED in SAS v. 9.2 (SAS Institute Inc., 2010).

## 3. Results and discussion

### 3.1. OTC adsorption in the presence of Brij 35

Adsorption plays an important role in the many factors affecting the transport of chemicals in soil. To the best of the author's knowledge, no study has been conducted to investigate the sorption behavior of OTC in the presence of nonionic surfactant Brij 35. Therefore, the sorption experiment in the presence and absence of Brij 35 was carried out; the results demonstrate that the amount of OTC sorbed was higher in the absence of Brij 35; however, it decreased in the presence of surfactant at concentrations above CMC.

Fig. 2 shows the sorption isotherms of OTC in Bij35 solutions at concentrations (0, 0.02, 0.25, 2.5 and 5 g L<sup>-1</sup>). All isotherms obtained for OTC fit well the Freundlich model (Eq. (1)), with R<sup>2</sup> values between 0.93 and 0.98. At low concentrations, sorption tends to be high which indicates the affinity of OTC to sorb to the soil; however, as the concentration increased, the sorption sites became saturated and prevented more OTC from being sorbed. Furthermore, *n* values which are presented in Table 3 indicated a Freundlich constant *n* < 1 which confirms the observation of saturated sites and the more OTC added, the more difficult to find available sites for binding. This is due to the presence of weak sorption energy. In terms of the value of the Freundlich sorption coefficient in the presence of Brij 35, *K<sub>f</sub>* shows the highest value when the surfactant is present as monomers (below CMC at 0.02 g L<sup>-1</sup>). As a result of increasing the concentration of Brij 35 to 0.25 g L<sup>-1</sup> (about 2.5 CMC), the *K<sub>f</sub>* value decreased and by increasing the surfactant concentration up to 5 g L<sup>-1</sup>, the continuous decrease of *K<sub>f</sub>* value is observed. This can be explained by the fact that at this high concentration, the surfactant is present in the solution in the micellar form which may facilitate the release of sorbed OTC into the solution.

The distribution coefficient (*K<sub>oc</sub>*) values have been used as an indicator of the leaching potential of herbicides such as atrazine and metribuzin (Southwick et al., 1995). The calculated *K<sub>oc</sub>* values show a higher leaching potential for OTC in the presence of Brij 35 (417.61–698.86) compared to the value of 844.10 in the absence of Brij 35.

Studies indicated that OTC has a great affinity to sorb to soil and clay (Sithole and Guy, 1987; Rabolle and Spliid, 2000); besides, it has a strong ability to chelate with monovalent and multivalent cations in soil organic matter and inorganic minerals by forming metal complexes which are considered the main reason for the immobility of OTC in soil (Martin, 1979; Figueroa and Mackay, 2005). However, under certain pH conditions, tetracyclines can be released from soil (Sithole and Guy, 1987; MacKay and Canterbury, 2005). The obtained *K<sub>f</sub>* values are found to be within the range reported by Jones et al. (2005). In the present study, *K<sub>f</sub>* values were found to be 23.55 in the absence of Brij 35. However, *K<sub>f</sub>* values are reduced noticeably to reach 12.76 and 14.53 in the presence of Brij 35; this shows low sorption ability. It should be noted that *K<sub>f</sub>* value in the presence of Brij 35 below CMC (in the monomer form) is slightly higher than *K<sub>f</sub>* value in the water solution with no surfactant because the presence of monomers can increase the affinity of Brij 35 to sorb to soil and hence no solubilization effect was

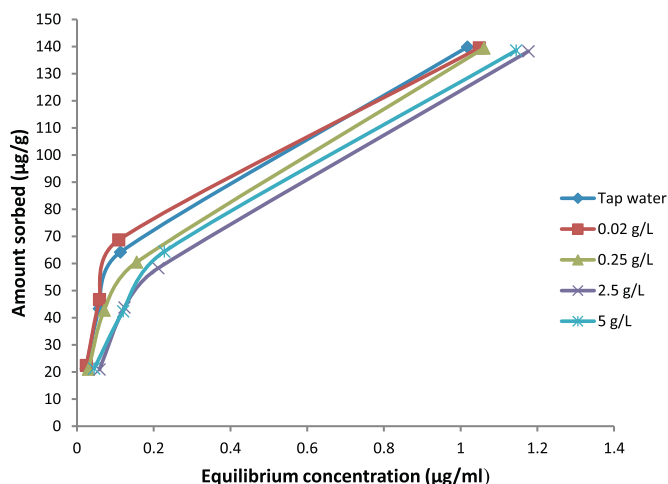


Fig. 2. Oxytetracycline isotherms in the absence (i.e. tap water) and presence of the non-ionic surfactant Brij 35 at concentrations below and above CMC.

Table 3

Distribution coefficient *K<sub>f</sub>*, *n*, R<sup>2</sup> and *K<sub>oc</sub>* values for OTC sorption in the presence of the nonionic surfactant, Brij 35.

Concentration of Brij 35 (g L <sup>-1</sup> )	<i>K<sub>f</sub></i>	<i>n</i>	R <sup>2</sup>	<i>K<sub>oc</sub></i>
0	23.55 ± 1.03	0.49 ± 0.02	0.95	844.10 ± 21.34
0.02	25.46 ± 1.54	0.45 ± 0.03	0.93	912.84 ± 31.96
0.25	19.49 ± 0.51	0.48 ± 0.01	0.97	698.86 ± 10.48
2.5	12.76 ± 0.30	0.60 ± 0.01	0.96	457.61 ± 6.28
5	14.53 ± 0.34	0.57 ± 0.01	0.98	520.82 ± 6.88

observed. Given the fact that surfactants are amphiphilic compounds, they sorb to soil after application. The sorption of nonionic surfactants onto soil has been studied (Kuhnt, 1993; Beigel et al., 1998; Shen and Yen, 1999; Shen, 2000; Paria, 2008).

Soil analysis showed that the zero point of charge (zpc) was achieved at pH about 3.4 and since the pH of our soil solution is 5.7, it is expected that soil surface would be negatively charged. Consequently, the dominant species of OTC is the zwitterionic OTC<sup>0+</sup> (at pH 3.3–7.3). This leads to increase in OTC sorption due to either electrostatic attraction or hydrogen bonding with soil surface (Sun et al., 2012). This is consistent with our results that showed high sorption coefficient of OTC in aqueous medium compared to the soil–water surfactant system, and confirm the relatively low mobility of OTC in soil (Log *P* = -1.12), as indicated by Herbert and Dorsey (1995).

However, in the presence of the non-ionic surfactant (above CMC), OTC exhibited low affinity to sorb to soil. This can be explained by the fact that above zpc, more donor sites are available. Therefore, hydrogen bonding interactions are promoted between the surfactant micelles and the soil surface resulting in a competition between Brij 35 and OTC for sorption sites. Additionally, the formation of micelles can further enhance OTC release in the solution. Therefore, in the presence of the surfactant monomers, greater sorption was observed.

As a matter of fact, nonionic surfactants at concentrations above CMC decreased the sorption coefficient (*K<sub>d</sub>*) values as the surfactants micelles in the aqueous phase compete with soil sorbed surfactants; they then become an effective partitioning medium for hydrophobic organic compounds (HOCs) leading to decreasing in *K<sub>d</sub>* values (Sun et al., 1995). Similarly, the effect of Brij 35 on the adsorption of atrazin was found to be proportional to the concentration. Since Brij 35 competes with atrazin for sorption sites in the

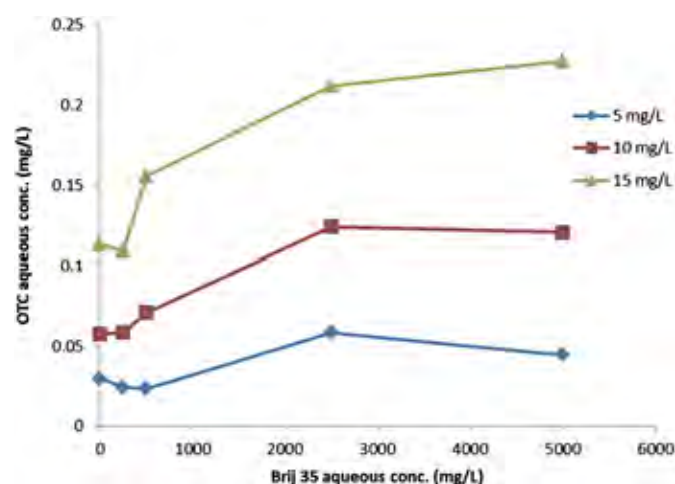


Fig. 3. Dependence of the aqueous concentration of OTC on the total concentration of the non-ionic surfactant (i.e. monomers and micelles).

soil, increasing the concentration will enhance the admicelles on the clay surface and in turn, dissolve atrazin (Chappell et al., 2005). The mobility of OTC in the water–soil–surfactant system is represented by Fig. 3. Generally, the aqueous concentration of OTC increased slightly below CMC and subsequently, the concentrations exhibited a sharp increase once CMC was exceeded. To explain this behavior, it is known that monomers are present in the solution when the surfactant concentrations are below CMC so the surfactant favored to sorb onto soil in higher concentrations. This eventually led to more OTC sorbed to soil but once CMC was reached, micelles present in the solution then compete with sorbed surfactant for OTC. The more micelles in the aqueous phase, the more OTC partitioning into the micelles within the solution resulting in higher concentrations released from sorbed OTC. This result matches the findings of Sun et al. (1995) and Wang and Keller (2008).

### 3.2. Mass balance

The mass balance for OTC is presented in Table 4. For all irrigation treatments, OTC concentrations were found to be below the detection limit by day 90. In both the water and surfactant-irrigated treatments, the total amount of OTC detectable in the system (soil + leachate), declined from an initial (applied) level of about 35 mg to less than 5 mg after 63 days and four irrigations. The relative amount of OTC in the two lower soil layers (15–45 and 45–70 cm) to that in the upper layers (0–5 and 5–15 cm) was at least 2-fold greater in the presence of surfactants from Day 21 onwards, suggesting that a greater amount of OTC leached downwards in the presence (vs. absence) of the surfactant. This will be addressed further in the discussion of OTC residue concentrations in the soil which follows.

From 21 days onward, the total quantity of detectable OTC was consistently greater in the surfactant treated columns than in those receiving tap water (Table 4). After 60 days, the total remaining OTC was 1.5 and 4 fold greater in the 5 and 0.5 g L<sup>-1</sup> treatments, respectively, than in the tap water treatment. The percentage of unrecovered OTC (Fig. 4) was similar for the irrigation types up to 21 days except for the treatment of Brij 35 0.5 mg L<sup>-1</sup> which showed higher percentages (before dropping again after the second irrigation) compared to other treatments; however, later values were lesser in the presence of surfactant at 5 mg L<sup>-1</sup>. In order to explain that, we should point out that the presence of the nonionic surfactant Brij 35 resulted in greater recovery compared to the tap water treatment. Despite the fact that OTC is immobile in soil under 5 cm (Blackwell et al., 2009), in the present study, 0.76% of the applied OTC was recovered in the leachate of the lysimeters irrigated with tap water, and 6.47 and 3.86%, respectively, for the 0.5 and 5.0 g L<sup>-1</sup> surfactant solutions.

The treatment differences in terms of the amount of unrecovered OTC (Fig. 4) indicate the existence of factors, other than leaching, in the loss of OTC in soil. These factors could include sorption, microbial degradation (Yang et al., 2009) and photodecomposition (Zhang et al., 2012). The fact that less OTC was recovered in the presence of surfactant indicates the effect of the surfactant on the degradation of OTC in soil.

Studies have shown that the presence of surfactant can increase the bioavailability of pesticides and hence, accelerate their degradation in the soil (Bardi et al., 2000). Similarly; Zhu et al. (2010) found that Brij 30 is capable of increasing the bioavailability of PAHs in the soil by enhancing the total abundance of bacteria. Recently, Zhang and Zhu (2012) reported that Tween-80 enhanced the sorption of pyrene on the bacterial cells and promoted degradation. However, in the present study, the presence of surfactant reduced, rather than enhanced, the loss of OTC (with lesser

**Table 4**  
Amounts of oxytetracycline (mg) in different soil profile depth ranges, cumulated in leachate, and overall, over a 63 day period (–, OTC was below detection levels).

Day	Irrigation solution		Soil profile depth range (cm)/(Cumulative leachate/Total ± S.D.)																	
	Tap water		Brij 35 @ 0.5 g L <sup>-1</sup>				Brij 35 @ 5 g L <sup>-1</sup>				Total									
	0–5	5–15	15–45	45–70	Cum. leachate	Total	0–5	5–15	15–45	45–70	Cum. leachate	Total	0–5	5–15	15–45	45–70	Cum. leachate	Total		
0	32.3	Irrigation I	–	–	–	–	32.25 ± 0.63	35.5	Irrigation I	–	–	–	–	35.52 ± 1.20	34.6	Irrigation I	–	–	–	34.56 ± 0.92
1	29.8	0.24	–	–	–	30.04 ± 0.38	25.1	4.23	–	–	–	–	29.61 ± 0.88	21.5	7.02	–	–	–	28.98 ± 0.59	
5	20.7	–	–	–	–	20.71 ± 0.50	13.4	8.04	–	–	–	–	21.75 ± 0.88	9.47	10.4	–	–	–	20.29 ± 0.45	
15	9.25	6.05	–	–	–	15.30 ± 0.20	4.64	1.44	5.77	–	–	–	12.16 ± 0.58	4.71	4.6	2.95	–	–	12.67 ± 0.56	
21	Irrigation II	–	–	–	–	15.30 ± 0.49	Irrigation II	–	–	–	–	–	13.58 ± 0.69	Irrigation II	–	–	–	–	13.57 ± 0.63	
22	4.09	0.75	2.25	–	–	7.09 ± 0.31	2.54	0.6	1.82	1.8	1.42	1.42	8.19 ± 0.29	2.28	0.17	1.21	0.69	5.26 ± 0.46		
30	1.84	0.5	0.35	–	–	2.69 ± 0.15	1.42	–	0.09	1.04	1.42	1.42	3.98 ± 0.17	0.66	–	–	1.44	3.00 ± 0.13		
42	Irrigation III	–	–	–	–	2.93 ± 0.19	Irrigation III	–	–	–	–	–	6.25 ± 0.39	Irrigation III	–	–	–	–	4.32 ± 0.30	
60	0.49	0.2	0.41	–	–	1.34 ± 0.08	–	–	–	0.3	2.27	2.27	2.57 ± 0.13	–	–	–	–	1.60 ± 0.13		
63	Irrigation IV	–	–	–	–	1.10 ± 0.04	Irrigation IV	–	–	–	–	–	4.87 ± 0.29	Irrigation IV	–	–	–	–	2.92 ± 0.19	

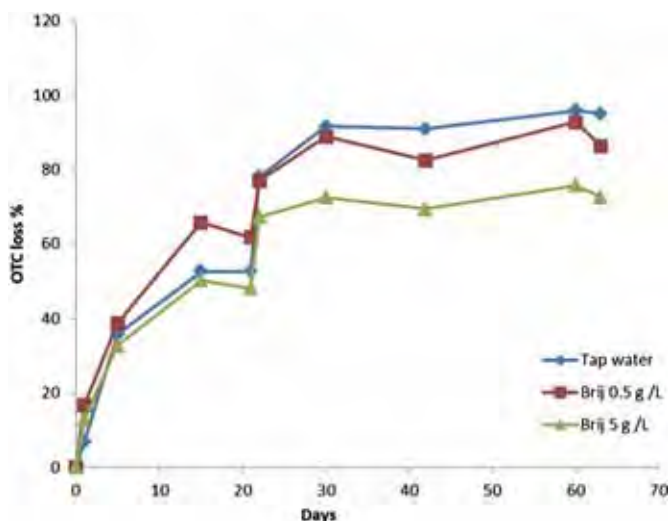


Fig. 4. OTC lost percentage in the season in the absence and presence of surfactant.

unrecovered amounts in the treatment of Brij 5 g L<sup>-1</sup>). This is consistent with the findings of Yoshioka and Stella (2002); they studied the effect of surfactants on the degradation of Beta-Lactam antibiotics and reported that inhibition of acid degradation of propicillin was observed in the presence of non-ionic surfactant Brij 35 (above CMC). On the contrary, anionic surfactants promoted degradation. Furthermore, the more hydrophobic the surfactant, the greater its inhibition effect on the antibiotic's breakdown. Recently, Mohammed-Ali (2012) reported that the non-ionic surfactant Tween-80 increased the stability of tetracycline solution. He indicated that, as a result of the formation of micelles, tetracycline molecules were captured by the micelles which prevented them from degradation to anhydrotetracycline.

### 3.3. Effect of Brij 35 on the OTC residues in soil

The downward movement of OTC in sandy soil was assessed in the presence of Brij 35 in the irrigation water and compared with tap water irrigation. Figs. 5–8 show the average concentrations of OTC in soil over the ninety day period for top soil 0.1, 0.3 and 0.6 m soil depth. The concentration of OTC exhibits different trends by depth.

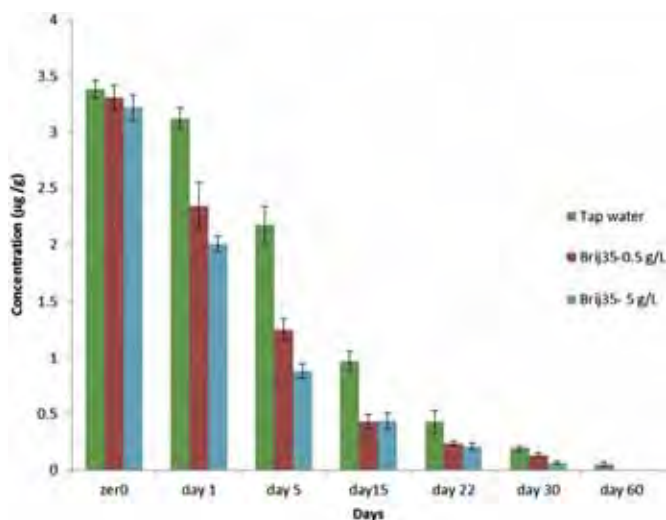


Fig. 5. Oxytetracycline concentration in the soil at the surface.

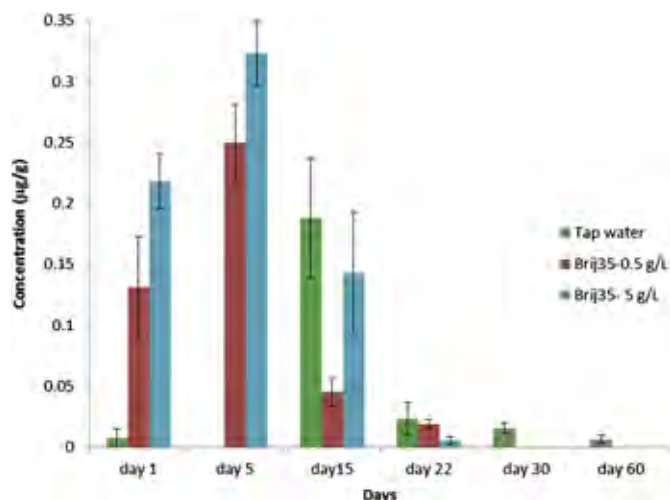


Fig. 6. Oxytetracycline concentration in the soil at 0.1 m depth.

As expected, the concentration at the soil surface was the highest in all treatments then decreased at greater depths. At the soil surface, the concentration decreased as the degradation rate increased by time; the concentration right after application was 3.38, 3.30 and 3.21 µg g<sup>-1</sup> in the tap water, Brij 35 0.5 and Brij 35 5 g L<sup>-1</sup>, respectively. At day 15, the concentration decreased significantly in the tap water (0.96 µg g<sup>-1</sup>) when compared to both treatments with surfactant (0.43 and 0.44 µg g<sup>-1</sup>). The results showed that OTC degraded over time to reach 0.05 µg g<sup>-1</sup> for tap water at day 60 whereas OTC was not detected in the treatments and by day 90, the OTC concentration was under the detection limit in samples taken from the tap water as well as for the surfactant treatments.

At 0.1 m depth, OTC concentrations in the surfactant treatments had an upward trend during the first 11 days followed by a sharp decrease; however, concentrations in the tap water increased from day 22 and then, decreased gradually to be non-detectable by day 90. At 0.3 m depth, the trend was similar to the 0.1 depth in all three treatments. This means that even with the tap water, OTC is still detected in soil samples and this is in consistent with Hamscher et al. (2002). They carried out a field study to investigate the fate of tetracyclines in soil fertilized by liquid manure and found that the detected concentrations of TCs increased by depth from

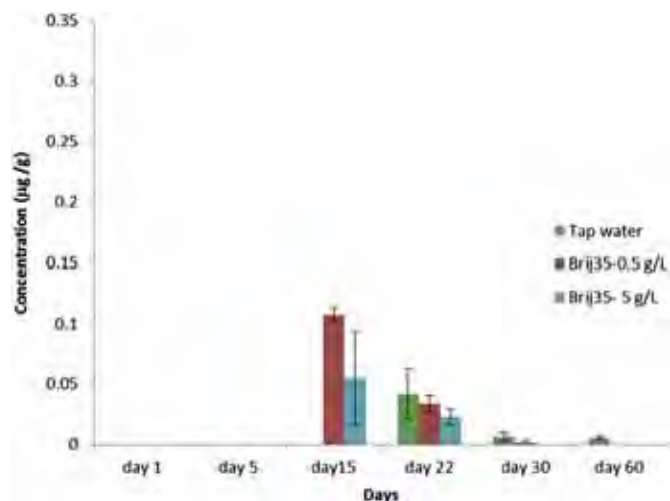


Fig. 7. Oxytetracycline concentration in the soil at 0.3 m depth.



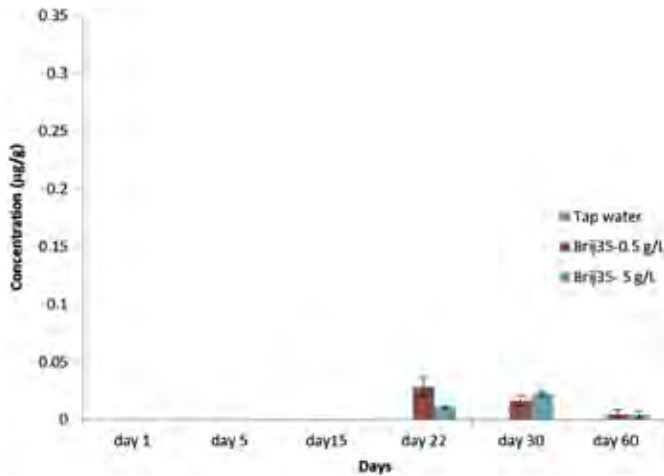


Fig. 8. Oxytetracycline concentration in the soil at 0.6 m depth.

171.7  $\mu\text{g kg}^{-1}$  in the 20–30 cm layer comparing to 86.2  $\mu\text{g kg}^{-1}$  in the top soil (0–10 cm). However, at the 0.6 m depth, considerable concentrations (20  $\mu\text{g kg}^{-1}$ ) of OTC were detected in the surfactant treatment only by day 22 then OTC concentrations decreased gradually in the surfactant treatment to (4.43  $\mu\text{g kg}^{-1}$ ) by day 60 while in tap water treatment, concentrations were under the detection limit. Moreover, the downward movement of OTC was significant in the treatment with Brij 35 0.5  $\text{g L}^{-1}$  compared to the treatment of Brij 35 5  $\text{g L}^{-1}$ . The decrease in the concentrations of OTC by time and depth indicates that OTC moved downwards in soil while at the same time it degraded over the ninety day period. Such a decrease is reflected as the significant effect ( $P \leq 0.01$ ) of treatment, depth, time and interaction among these factors (Table 5).

Given that OTC is strongly sorbed to soils, the enhanced mobility of OTC in soil in the presence of surfactant as compared to tap water may be due to the interaction between soil and surfactant. As indicated earlier, the sorption of nonionic surfactants onto soil is an important factor and may result in competition between Brij 35 and OTC for binding sites in the soil organic matter. This assumption is in agreement with the findings of Chefetz et al. (2008) who studied the effect of irrigation with secondary treated wastewater (STWW) on the environmental behavior of pharmaceutical compounds in soil columns and reported that naproxen was mobile in the soil irrigated with STWW. This behavior was the result of competition between DOM which present in the wastewater and naproxen for the binding sites. A similar conclusion was obtained by Kulshrestha et al. (2004) who stated that dissolved natural organic acids, such as humic acids, may compete with clay for oxytetracycline and thus, enhance its mobility in the soil profile.

In general, surfactants can increase the apparent water solubility of contaminants through their solubilization in the micelles,

**Table 5**  
Repeated measures analysis of variance for oxytetracycline residues in soil.

Effect	Probability
Treatment	<0.0001**
Depth	<0.0001**
Treatment $\times$ Depth	<0.0001**
Time	<0.0001**
Treatment $\times$ Time	0.2031 <sup>NS</sup>
Depth $\times$ Time	<0.0001**
Treatment $\times$ Depth $\times$ Time	<0.0001**

\*\*denotes significant difference ( $P \leq 0.01$ ).

<sup>NS</sup>denotes insignificant difference ( $P \leq 0.01$ ).

which form at concentrations above their CMC (Edwards et al., 1991). Depending on this concentration, contaminant properties, the degree and the location of solubilization in the micelle, surfactant effects can vary. Therefore, in our study, the concentration of Brij 35 above CMC has the potential to form micelles which in turn, have the ability to react with soil organic matter (SOM) by sorption as well as increasing OTC solubilization in the aqueous solution. The observed downward movement of OTC in the lysimeters from the surface to 0.6 m depth can be explained by the fact that Brij 35 is competing efficiently with OTC for sorption sites, while above the CMC, the lipophilic end (alkyl chain) of the surfactant molecule aggregate together inside the micelle, with the hydrophilic end point toward the aqueous phase on the exterior (Cheah et al., 1998). It is assumed that the polar moiety of Brij 35 (the exterior) could interact with the negatively charged sites in soil thus occupying the sites that otherwise would be available for interaction with the functional groups of OTC. Due to this behavior, OTC would exhibit weak sorption energy with soil. Another reason that may contribute to the Brij 35 competition for soil sites is the huge molecular size (1199.55) compared to OTC (284.72) (Yong and Mulligan, 2004). Furthermore, Muller et al. (2007) reported that wastewater irrigation was found to enhance the mobility of contaminants such as pesticides. They also stated that the presence of surfactants in the irrigation water is considered as a key factor in the promotion of the leaching potential of pesticides. Additionally, non-ionic alcohol ethoxylates (AEs) are found to enhance the mobility of pesticides when applied at high (vs. low) concentrations to soil columns (Huggenberger et al., 1973). On the other hand, studies have shown that repeated applications of manure on soil as a fertilizer not only can lead to the release of bound tetracyclines from the soil and consequently increase its downward movement through soil (Kulshrestha et al., 2004) but also it poses a risk factor for public health as it may enhance the transfer of antibacterial-resistant bacteria to humans through water or plants (Sengelov et al., 2003).

#### 3.4. Effect of surfactant on OTC residues in water

The leachate samples were analyzed to quantify the amount of oxytetracycline leached in the shallow ground water. OTC concentrations were higher in the leachate in the both treatments with Brij 35 ranged from (102.72 – 2.72) and (49.66 – below detection limit)  $\mu\text{g L}^{-1}$  for Brij 35 0.5 and Brij 35 5  $\text{g L}^{-1}$ , respectively

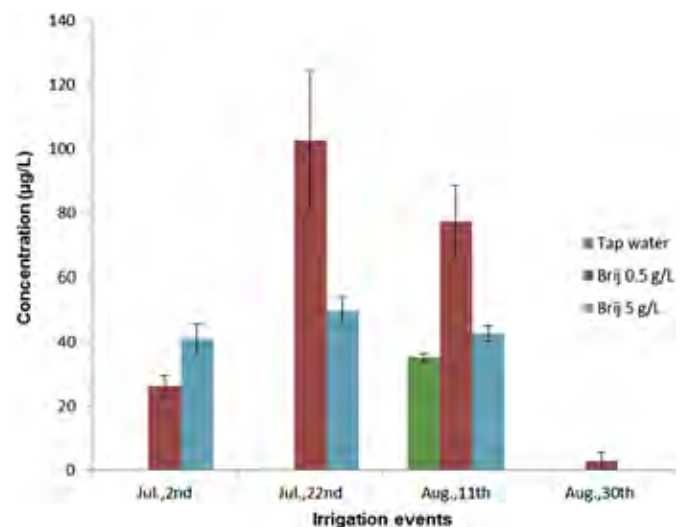


Fig. 9. Oxytetracycline concentration in the leachate.



(Fig. 9). In addition, statistical analysis indicated that a considerable concentration of OTC in the treatment with Brij 35  $0.5 \text{ g L}^{-1}$  was observed when compared to the treatment with Brij 35  $5 \text{ g L}^{-1}$ . On the other hand, the OTC concentration was non-detectable in the leachate of the tap water except for the third application (38 days after the beginning of the season) OTC concentration was  $35.23 \mu\text{g L}^{-1}$ ; this may have leached into the ground water through macrospores, generated by earthworms and the plant root system. In the present study, the maximum concentration in the leachate ( $102.72 \mu\text{g L}^{-1}$ ) was higher than the concentration of  $72 \mu\text{g L}^{-1}$  which was detected in the surface water near the plots were treated with OTC-contaminated manure (Kay et al., 2005).

The presence of OTC in the leachate in the treatment with both concentrations of Brij 35 may have two possible explanations. First, Brij 35 has the ability to form micelles which dissolve OTC and increase its mobility after each irrigation event. Second, the effect of surfactant on soil properties; several studies have been conducted in this area of research (Koren, 1972; Abu-Zreig et al., 1999; Singh and Kumar, 2000). They concluded that nonionic surfactants increased the water movement by increasing soil dispersion. Similar findings were reported by Tao et al. (2006) who confirmed that soil dispersion as a result of the presence of surfactants, increased the sorption sites and played an important role in contaminants movement. Furthermore, nonionic surfactants enhance the infiltration by increasing water penetration due to their ability to increase soil dispersion (Abu-Zreig et al., 2003; Katagi, 2008).

Since nonionic surfactants are known for their ability to release pesticides from aged soils, Mata-Sandoval et al. (2000) reported that Triton X-100 showed good desorption efficiency of pesticides from soil by replacing the pesticide on the sorption sites on soil particles and releasing it from soil. Xiarchos and Doulia (2006) stated that the ability of solubilizing polar compounds is positively linked to the hydrophilicity, and this hydrophilicity is known to be governed by the number of ethylene oxide groups (EO) in the molecule, i.e., the greater the number of EO groups, the more hydrophilic the surfactant (Jafvert et al., 1994). Since Brij 35 has twenty-three EO groups, it is expected to exhibit a high capacity to solubilize polar compounds. In terms of OTC structure, Hayes et al. (2007) indicated that OTC contains polar functional groups such as the cationic amide and ortho-substituted anionic groups which play an important role in the binding process of OTC and soil. When the cationic amide group attracts to the soil surface it increases the closeness of the ortho anionic moiety to the negatively charged surface which cause repulsion and hence decrease OTC sorption. They also found that the constant of binding is higher when OTC interacts with an aromatic acid vs. an aliphatic acid. This supports the results obtained by our study. Because Brij 35 has a long aliphatic structure represented by the ethylene oxide group  $(\text{CH}_2\text{CH}_2\text{O})_{23}$  which may explain the leaching behavior of OTC in this study.

#### 4. Conclusion

Understanding the mobility of Oxytetracycline within the soil–water–surfactant system is an important step in assessing the risk resulting from irrigation with wastewater. The present study showed that the presence of nonionic surfactant (Brij 35) in concentrations higher than its CMC, increased the downward movement of OTC in the soil profile, presumably due to the solubilization power of micelles in the soil–water system when Brij 35 was present at  $0.5$  or  $5 \text{ g L}^{-1}$ . In addition, the presence of Brij 35 significantly increased the leaching potential. These findings were supported by the sorption test which indicated that  $K_f$  values for OTC were significantly lower in the presence of the nonionic surfactant. Despite the fact that lysimeter studies have shown OTC to seldom move beyond  $10 \text{ cm}$  depth in the soil profile, our study

indicates that the presence of a nonionic surfactant in irrigation water promoted the mobility of OTC in soil and led to its presence in leachate at a depth of  $1 \text{ m}$  below the soil surface. Therefore, the continuous use of wastewater containing nonionic surfactants at concentrations exceeding their CMC can be expected to increase the likelihood of OTC being found in leachate, eventually reaching ground water, and potentially contaminating water resources. This study highlights the role played by the application of OTC-contaminated manure as is a common practice; it does not only increase the risk of OTC leaching but also results in the long term exposure of bacteria to antibiotics. This increases the risk of the bacteria becoming resistant to antibiotics as well as increasing the likelihood of transferring the bacteria resistance to humans. This could become a momentous issue, as many antibiotics, including OTC, remain biologically active in the soil for weeks or months.

#### Acknowledgment

The authors would like to acknowledge the financial support provided by Faculty of Agriculture, Zagazig University, Egypt and the Natural Sciences and Engineering Research Council of Canada (NSERC).

#### References

- Abu-Zreig, M., Rudra, R.P., Dickinson, W.T., Evans, L.J., 1999. Effect of surfactants on sorption of atrazine by soil. *J. Contam. Hydrol.* 36, 249–263.
- Abu-Zreig, M., Rudra, R.P., Dickinson, W.T., 2003. Effect of application of surfactants on hydraulic properties of soils. *Biosyst. Eng.* 84 (3), 363–372.
- Arikan, O., Sikora, L., Mulbry, W., Khan, S., Foster, G., 2007. Composting rapidly reduces levels of extractable oxytetracycline in manure from therapeutically treated beef calves. *Bioresour. Technol.* 98, 169–176.
- Aronstein, B.N., Calvillo, Y.M., Alexander, M., 1991. Effect of surfactants at low concentrations on the desorption and biodegradation of sorbed aromatic compounds in soil. *Environ. Sci. Technol.* 25, 1728–1731.
- Barber, L.B., Keefe, S.H., Leblanc, D.R., Bradley, P.M., Chappelle, F.H., Meyer, M.T., Loftin, K.A., Kolpin, D.W., Rubio, F., 2009. Fate of sulfamethoxazole, 4-nonylphenol, and 17-estradiol in groundwater contaminated by wastewater treatment plant effluent. *Environ. Sci. Technol.* 43, 4843–4850.
- Bardi, L., Mattei, A., Steffan, S., Marzona, M., 2000. Hydrocarbon degradation by a soil microbial population with  $\beta$ -cyclodextrin as surfactant to enhance bioavailability. *Enzyme Microb. Technol.* 27, 709–713.
- Beigel, C., Barriuso, E., Calvet, R., 1998. Sorption of low levels of nonionic and anionic surfactants on soil: effects on sorption of triticonazole fungicide. *Pestic. Sci.* 54, 52–60.
- Blackwell, P.A., Kay, P., Boxall, A.B.A., 2007. The dissipation and transport of veterinary antibiotics in a sandy loam soil. *Chemosphere* 67, 292–299.
- Blackwell, P.A., Kay, P., Ashauer, R., Boxall, A.B.A., 2009. Effects of agricultural conditions on the leaching behaviour of veterinary antibiotics in soils. *Chemosphere* 75, 13–19.
- Boxall, A.B.A., Johnson, P., Smith, E.J., Sinclair, C.J., Stutt, E., Levy, L., 2006. Uptake of veterinary medicines from soils into plants. *J. Agric. Food Chem.* 54, 2288–2297.
- Brunner, P.H., Capri, S., Marcomini, A., Giger, W., 1988. Occurrence and behaviour of linear alkylbenzenesulphonates, nonylphenol, nonylphenol mono- and nonylphenol diethoxylates in sewage and sewage sludge treatment. *Water Res.* 22 (12), 1465–1472.
- Castillo, M., Alonso, M.C., Riu, J., Barceloa, D., 1999. Identification of polar, ionic, and highly water soluble organic pollutants in untreated industrial wastewaters. *Environ. Sci. Technol.* 33, 1300–1306.
- Chappell, M.A., Laird, D.A., Thompson, M.L., Evangelou, V.P., 2005. Cosorption of atrazine and a lauryl polyoxyethylene oxide nonionic surfactant on smectite. *J. Agric. Food Chem.* 53, 10127–10133.
- Chatterjee, R., 2008. Fresh produce from wastewater. *Environ. Sci. Technol.* 42 (21), 7732.
- Cheah, P.S., Reible, D., Valsaraj, K.T., Constant, D., Walsh, W., Thibodeaux, L.J., 1998. Simulation of soil washing with surfactants. *J. Hazard. Mater.* 59, 107–122.
- Chefetz, B., Muallem, T., Ben-Ari, J., 2008. Sorption and mobility of soils irrigated with reclaimed wastewater. *Chemosphere* 73, 1335–1343.
- Chettri, M., Thapa, U., 2004. Integrated nutrient management with farm yard manure on potatoes (*Solanum tuberosum*) under gangetic plains of West Bengal. *Environ. Ecol.* 22, 766–769.
- Conn, K.E., Barber, L.B., Brown, G.K., Siegrist, R.L., 2006. Occurrence and fate of organic contaminants during onsite wastewater treatment. *Environ. Sci. Technol.* 40, 7358–7366.
- Drechsel, C., Raschid-Sally, L., Redwood, M., Bahri, A., 2010. Water Irrigation and Health. Assessing and Mitigating Risk in Low-income Countries. Earthscan/IDRC, UK, pp. 432.

- Edwards, D.A., Luthy, R.G., Liu, Z., 1991. Solubilization of polycyclic aromatic hydrocarbons in micellar nonionic surfactant solutions. *Environ. Sci. Technol.* 25, 127–133.
- Esperanza, M., Suidan, M.T., Nishimura, F., Wang, Z., Sorial, G.A., 2004. Determination of sex hormones and nonylphenol ethoxylates in the aqueous matrixes of two pilot-scale municipal wastewater treatment plants. *Environ. Sci. Technol.* 38, 3028–3035.
- Field, J.A., Leenheer, J.A., Thorn, K.A., Barber, L.B., Rostad, C., Macalady, D.L., Daniel, S.R., 1992. Identification of persistent anionic surfactant derived chemicals in sewage effluent and ground water. *J. Contam. Hydrol.* 9, 55–78.
- Figuerola, R.A., Mackay, A., 2005. Sorption of oxytetracycline to iron oxides and iron oxide-rich soils. *Environ. Sci. Technol.* 39, 6664–6671.
- Halling-Sorensen, B., Nielsen, S.N., Lanzky, P.F., Ingerslev, F., Lutzhoft, H.C.H., Jorgensen, S.E., 1998. Occurrence, fate and effects of pharmaceutical substances in the environment – a review. *Chemosphere* 36 (2), 357–393.
- Hamilton, A.J., Stagnitti, F., Xiong, X., Kreidl, S.L., Benke, K.K., Maher, P., 2007. Wastewater irrigation: the state of play. *Vadose Zone J.* 6, 823–840.
- Hamscher, G., Abu-Quare, A., Sczesny, S., Hoper, H., Nau, H., 2000. Determination of tetracyclines and tylosin in soil and water samples from agricultural areas in lower saxony. In: van Ginke, L.A., Ruiter, A. (Eds.), *Proceedings of the Euroside IV Conference, Veldhoven, Netherlands, 8–10 May 2000*. National Institute of Public Health and the Environment (RIVM), Bilthoven, Netherlands, pp. 522–526.
- Hamscher, G., Sczesny, S., Hoper, H., Nau, H., 2002. Determination of persistent tetracycline residues in soil fertilized with liquid manure by high-performance liquid chromatography with electro-spray ionization tandem mass spectrometry. *Anal. Chem.* 74, 1509–1518.
- Hayes, P.L., Gibbs-Davis, J.M., Musorrafiti, M.J., Mifflin, A.L., Scheidt, K.A., Geiger, F.M., 2007. Environmental biogeochemistry studied by second-harmonic generation: a look at the agricultural antibiotic oxytetracycline. *J. Phys. Chem. C* 111, 8796–8804.
- HERA, 2009. *Human & Environmental Risk Assessment on Ingredients of European Household Cleaning Products. Alcohol Ethoxylates Report, Version 2.0*.
- Herbert, B.J., Dorsey, J.G., 1995. n-Octanol–water partition coefficient estimation by micellar electrokinetic capillary chromatography. *Anal. Chem.* 67, 744–749.
- Himmelsbach, M., Buchberger, W., 2005. Residue analysis of oxytetracycline in water and sediment samples by high performance liquid chromatography and immunochemical techniques. *Microchim. Acta* 151, 67–72.
- Huggenberger, F.H., Letey, J., Farmer, W.J., 1973. Effect of two non-ionic surfactants on adsorption and mobility of selected pesticides in a soil-system. *Soil Sci. Soc. Am. Proc.* 37, 215–219.
- Jafvert, C.T., van Hoof, P.L., Heath, J.K., 1994. Solubilization of non-polar compounds by non-ionic surfactant micelles. *Water Res.* 28, 1009–1017.
- Jones, A.D., Bruland, G.L., Agrawal, S.G., Vasudevan, D., 2005. Factors influencing the sorption of oxytetracycline to soils. *Environ. Toxicol. Chem.* 24 (4), 761–770.
- Kan, T.A., Tomson, M.B., 1990. Ground water transport of hydrophobic organic compounds in the presence of dissolved organic matter. *Environ. Toxicol. Chem.* 9, 253–263.
- Katagi, T., 2008. Surfactant effects on environmental behavior of pesticides. *Rev. Environ. Contam. Toxicol.*, 71–177.
- Kay, P., Blackwell, P.A., Boxall, A.B.A., 2005. Transport of veterinary antibiotics in overland flow following the application of slurry to arable land. *Chemosphere* 59 (7), 951–959.
- Koren, E., 1972. Leaching of trifluralin and oryzalin in soil with three surfactants. *Weed Sci.* 20, 230–232.
- Kuhnt, G., 1993. Behavior and fate of surfactants in soil. *Environ. Toxicol. Chem.* 12, 1813–1820.
- Kulshrestha, P., Giese, R.F.J., Aga, D.S., 2004. Investigating the molecular interactions of oxytetracycline in clay and organic matter: insights on factors affecting its mobility in soil. *Environ. Sci. Technol.* 38, 4097–4105.
- Loke, M.L., Tjornelund, J., Halling-Sorensen, B., 2002. Determination of the distribution coefficient (Log K<sub>d</sub>) of oxytetracycline, tylosin A, olaquinox and metronidazole in manure. *Chemosphere* 48, 351–361.
- Loyo-Rosales, J.E., Rice, C.P., Torrents, A., 2007. Fate of octyl and nonylphenol ethoxylates and some carboxylated derivatives in three American wastewater treatment plants. *Environ. Sci. Technol.* 41, 6815–6821.
- MacKay, A.A., Canterbury, B., 2005. Oxytetracycline sorption to organic matter by metal-bridging. *J. Environ. Qual.* 34, 1964–1971.
- Martin, S.R., 1979. Equilibrium and kinetic studies on the interaction of tetracyclines with calcium and magnesium. *Biophys. Chem.* 10, 319–326.
- Martinez-Carballo, E., Gonzalez-Barreiro, C., Scharf, S., Gans, O., 2007. Environmental monitoring study of selected veterinary antibiotics in animal manure and soils in Austria. *Environ. Pollut.* 148, 570–579.
- Mata-Sandoval, J.C., Karns, J., Torrents, A., 2000. Effects of rhamnolipids produced by *Pseudomonas aeruginosa* UG2 on the solubilization of pesticides. *Environ. Sci. Technol.* 34, 4923–4930.
- Mohammed-Ali, M.A.J., 2012. Stability study of tetracycline drug in acidic and alkaline solutions by colorimetric method. *J. Chem. Pharm. Res.* 4 (2), 1319–1326.
- Muller, K., Magesa, G.N., Bolan, N.S., 2007. A critical review of the influence of effluent irrigation on the fate of pesticides in soil. *Agric. Ecosyst. Environ.* 120, 93–116.
- Narkis, N., Ben-David, B., 1985. Adsorption of non-ionic surfactants on active carbon and mineral clay. *Water Res.* 19, 815–824.
- Nilufer, F., 2005. Fate and transport of herbicides in soil in the presence of surfactants in the irrigation water. M.Sc. thesis, McGill University, 138 pp. (un published).
- Paria, S., 2008. Surfactant-enhanced remediation of organic contaminated soil and water. *Adv. Colloid Interf. Sci.* 138, 24–58.
- PPDB, 2009. *The Pesticide Properties Database Report*. Agriculture and Environment Research Unit. Science & Technology Research Institute, University of Hertfordshire, College Lane, Hatfield, Herts, AL10 9AB, UK.
- Rabolle, M., Spliid, N., 2000. Sorption and mobility of metronidazole, olaquinox, oxytetracycline and tylosin in soil. *Chemosphere* 40, 715–722.
- Rhodes, G., Huys, G., Swings, J., Mccann, P., Hiney, M., Smith, P., Pickup, R., September 2000. Distribution of oxytetracycline resistance plasmids between aeromonads in hospital and aquaculture environments: implication of Tn1721 in dissemination of the tetracycline resistance determinant tet A. *Appl. Environ. Microbiol.* 66, 3883–3890. <http://dx.doi.org/10.1128/AEM.66.9.3883-3890.2000>.
- Rodriguez-Cruz, M.S., Sanchez-Martin, M.J., Sanchez-Camazano, M., 2004. Enhanced desorption of herbicides sorbed on soils by addition of Triton X-100. *J. Environ. Qual.* 33, 920–929.
- Sanchez-Camazano, M., Arienzo, M., Sanchez-Martin, M.J., Crisanto, T., 1995. Effect of different surfactants on the mobility of selected non-ionic pesticides in soil. *Chemosphere* 31, 3793–3801.
- SAS Institute Inc., 2010. *SAS/GRAPH®9.2 Reference*, second ed. SAS Institute Inc., Cary, NC.
- Sassman, S.A., Lee, L.S., 2005. Sorption of three tetracyclines by several soils: assessing the role of pH and cation exchange. *Environ. Sci. Technol.* 39, 7452–7459.
- Sengelov, G., Agerso, Y., Halling-Sorensen, B., Baloda, S.B., Andersen, J.S., Jensen, L.B., 2003. Bacterial antibiotic resistance levels in Danish farmland as a result of treatment with pig manure slurry. *Environ. Int.* 28, 587–595.
- Shen, Y.H., 2000. Sorption of non-ionic surfactants to soil: the role of soil mineral composition. *Chemosphere* 41 (5), 711–716.
- Shen, Y.H., Yen, M.H., 1999. Sorption of nonionic surfactants on soil. *Environ. Technol.* 20, 425–430.
- Singh, R.P., Kumar, R., 2000. Evaluation of the effect of surfactants on the movement of pesticides in soils using a soil thin-layer chromatography technique. *Soil Sedim. Contam.* 9, 407–423.
- Sithole, B.B., Guy, R.G., 1987. Models for tetracycline in aquatic environment. I. Interaction with bentonite clay systems. *Water Air Soil Pollut.* 32, 303–314.
- Southwick, L.M., Willis, G.H., Johnson, D.C., Selim, H.M., 1995. Leaching of nitrate, atrazine and metribuzin from sugarcane in southern Louisiana. *J. Environ. Qual.* 24, 684–690.
- Stuart, M., Lapworth, D., Crane, E., Hart, A., 2012. Review of risk from potential emerging contaminants in UK groundwater. *Sci. Total Environ.* 416, 1–21.
- Sun, S., Inskip, W.P., Boyd, S., 1995. Sorption of nonionic compounds in soil–water systems containing a micelle-forming surfactant. *Environ. Sci. Technol.* 29 (4), 903–913.
- Sun, Y., Yue, Q., Gao, B., Li, Q., Huang, L., Yao, F., Xu, X., 2012. Preparation of activated carbon derived from cotton linter fibers by fused NaOH activation and its application for oxytetracycline (OTC) adsorption. *J. Colloid Interf. Sci.* 368, 521–527.
- Tao, Q., Dong, S., Hong, X., Wang, T., 2006. Effect of surfactants at low concentrations on the sorption of atrazine by natural sediment. *Water Environ. Res.* 78, 653–660.
- Terlaak, T.L., Gebbink, W.A., Tolls, J., 2006. The effect of pH and ionic strength on the sorption of sulfachloropyridazine, tylosin and oxytetracycline to soil. *Environ. Toxicol. Chem.* 25 (4), 904–911.
- Thiele Bruhn, S., 2003. Pharmaceutical antibiotic compounds in soils – a review. *J. Plant Nutr. Soil Sci.* 166, 145–167.
- Uslu, M.O., Balcioglu, I.A., 2009. Simultaneous removal of oxytetracycline and sulfamethazine antibacterials from animal waste by chemical oxidation processes. *J. Agric. Food Chem.* 57, 11284–11291.
- Wang, P., Keller, A.A., 2008. Partitioning of hydrophobic organic compounds within soil–water–surfactant systems. *Water Res.* 42, 2093–2101.
- Wang, Q., Yates, S.R., 2008. Laboratory Study of oxytetracycline degradation kinetics in animal manure and soil. *J. Agric. Food Chem.* 56, 1683–1688.
- Wild, S.B., Waterrath, K.S., Jones, K.F., 1990. Organic contaminants in an agricultural soil with a known history of sewage sludge amendments. *Environ. Sci. Technol.* 24 (11), 1706–1711.
- Wilde, D.T., Mertens, J., Spanoghe, P., Ryckeboer, J., Jaeken, P., Springael, D., 2008. Sorption kinetics and its effects on retention and leaching. *Chemosphere* 72, 509–516.
- Xiarchos, I., Douliá, D., 2006. Effect of nonionic surfactants on the solubilization of alachlor. *J. Hazard. Mater.* B136, 882–888.
- Yang, J.F., Ying, G.G., Zhou, L.J., Liu, S., Zhao, J.L., 2009. Dissipation of oxytetracycline in soils under different redox conditions. *Environ. Pollut.* 157, 2704–2709.
- Yong, R.N., Mulligan, C.N., 2004. *Natural Attenuation of Contaminants in Soil*. CRC Press, Lewis Publishers, USA, pp. 319.
- Yoshioka, S., Stella, V., 2002. *Stability of Drugs and Dosage Forms*. Kluwer Academic Publishers, NY, pp. 268.
- Zhang, D., Zhu, L., 2012. Effects of Tween 80 on the removal, sorption and biodegradation of pyrene by *Klebsiella oxytoca* PYR-1. *Environ. Pollut.* 164, 169–174.
- Zhang, Y., Cai, X., Lang, X., Qiao, X., Li, X., Chen, J., 2012. Insights into aquatic toxicities of the antibiotics oxytetracycline and ciprofloxacin in the presence of metal: complexation versus mixture. *Environ. Pollut.* 166, 48–56.
- Zhu, H., Singleton, D.S., Aitken, M.D., 2010. Effects of nonionic surfactant addition on populations of polycyclic aromatic hydrocarbon-degrading bacteria in a bioreactor treating contaminated soil. *Environ. Sci. Technol.* 44, 7266–7271.

## Evaluation of WARMF model for flow and nitrogen transport in an agricultural watershed under a cold climate

Shadi Dayyani, Shiv O. Prasher, Ali Madani, Chandra A. Madramootoo and Peter Enright

### ABSTRACT

The Watershed Analysis Risk Management Framework (WARMF) model is adapted to simulate flow and nitrate-N transport in an agricultural watershed in Quebec, Canada. The model was evaluated for the St. Esprit Watershed (24.3 km<sup>2</sup>), which is a part of the 210 km<sup>2</sup> St. Esprit river basin, a tributary of the L'Assomption Watershed (4,220 km<sup>2</sup>). WARMF's hydrologic calibration and validation was performed using data from the gauge station located at the outlet of the watershed. Water-quality data collected were used to guide water quality calibration/validation. Simulations were carried out from 1994 to 1996; data from 1994 and 1995 were used for model calibration and data from 1996 were used for model validation. The model performed reasonably well in simulating the hydrologic response and nitrate losses at the outlet of the watershed. The  $R^2$  between the observed and simulated monthly stream flow for calibration was 0.92, and that for validation was 0.94. The corresponding coefficients of efficiency ( $E$ ) were 0.89 and 0.91. The  $R^2$  and  $E$  values for calibration/validation of NO<sub>3</sub>-N loads simulation were 0.89/0.84 and 0.86/0.75, respectively. Thus, the model simulated monthly flow and nitrogen losses with a good degree of accuracy over the entire year.

**Key words** | cold climate, hydrology, modeling, WARMF, water quality, watershed scale

### INTRODUCTION

The agricultural sector in Canada in general and Quebec in particular has witnessed substantial growth over the past two decades. The increase in agricultural production can be attributed to several factors, such as mechanization of farm operations, soil and water management, use of chemical fertilizers, and improved crop varieties (Gollamudi 2006). At the same time, this has placed the region's water bodies under severe environmental stress. In Quebec, agriculture is responsible for over 70% of the total non-point source pollution (Enright & Madramootoo 2004). Increased levels of phosphorus and nitrogen in lakes and rivers promote eutrophication, a phenomenon that may result in excessive growth of algae and poisonous cyanobacteria that deplete dissolved oxygen and render it hazardous for aquatic as well as human life. In Quebec, nitrogen and

phosphorus contamination of watercourses and lakes is largely attributed to non-point source pollution from agricultural fields (Gollamudi 2006).

While long-term field-scale monitoring is necessary to establish a theoretical understanding of nutrient dynamics, only a limited number of studies are available due to the high cost of instrumentation and operation (Gollamudi 2006). Additionally, collecting long-term data for a range of climatic, hydrologic, and topographic conditions is a time-consuming and difficult process. Thus, complementing real-time field data with a validated hydrological and water quality simulation model is both cost-effective and time-efficient.

Hydrological and water quality simulation models have developed from elementary to complex algorithms in the

**Shadi Dayyani** (corresponding author)

**Shiv O. Prasher**  
Department of Bioresource Engineering,  
McGill University, Montreal, Quebec,  
Canada

E-mail: [shadi.dayyanidardashti@mail.mcgill.ca](mailto:shadi.dayyanidardashti@mail.mcgill.ca)

**Ali Madani**

Department of Engineering,  
Nova Scotia Agricultural College,  
Truro, Nova Scotia,  
Canada

**Chandra A. Madramootoo**

Faculty of Agricultural and Environmental  
Sciences,  
McGill University, Montreal, Quebec,  
Canada

**Peter Enright**

Farm Management and Technology Program,  
Faculty of Agricultural and Environmental  
Sciences, McGill University,  
Montreal, Quebec,  
Canada

past three decades (Gollamudi 2006). A common starting point for all of these models is the necessity to accurately simulate the movement of water through different pathways of the hydrologic cycle – precipitation, overland flow, infiltration, subsurface flow, deep seepage, evapotranspiration (ET), and stream flow. The ability of a model to accurately simulate hydrological processes, such as surface runoff and subsurface drain flow, is important for reliable predictions of nutrient losses. The key criteria in choosing a model include the availability of reliable input data for the model parameters, spatial/temporal scale of use, and nature of output.

In this study, we used the Watershed Analysis Risk Management Framework (WARMF) model (Chen *et al.* 1998), to predict water quality and quantity in an agricultural watershed in Quebec, Canada. WARMF is a user-friendly tool, organized into five linked modules under one geographic information system (GIS)-based graphical user interface (GUI). It was developed under the sponsorship of the Electric Power Research Institute (EPRI) as a decision support system for watershed management. The scientific basis of the model has undergone several peer reviews by independent experts under US EPA guidelines (EPRI 2000). WARMF has been applied to over 15 watersheds in the United States and internationally (Chen *et al.* 2001a; Weintraub *et al.* 2001b, 2004; Herr *et al.* 2002; Keller *et al.* 2004; Geza & McCray 2007; Rambow *et al.* 2008). The focus of these studies has varied from total mass daily loading (TMDL) calculation (nutrients, sediment, fecal coliform, metals) to more research-oriented applications, such as modeling the fate and transport of mercury in a watershed and the impact of onsite wastewater systems on a watershed scale. The size of river basin applications ranges from 28 to 42,000 km<sup>2</sup>. There is no limit on the size or scale of a potential WARMF application as long as adequate topography data are available (USEPA 2009).

WARMF is primarily a surface hydrology model that, to the best of our knowledge, has not been tested under cold climatic conditions and on predominantly subsurface-drained agricultural watersheds. Therefore, this study provided an evaluation of the WARMF model for a small agricultural watershed in eastern Canada.

The main objective of this study was to evaluate the WARMF model to: (1) estimate flow; and (2) estimate nitrate-nitrogen losses in an agricultural watershed in a

cold region using three-site years of data. The evaluated model is a decision support system for watershed managers to perform TMDL analysis or to evaluate the impacts of different management practices to improve the quality of water. WARMF can be used to simulate the nitrogen load leaving a watershed, enabling managers to determine how load reductions can be most effectively allocated.

## MATERIALS AND METHODS

### WARMF model description

WARMF (Chen *et al.* 1998) is classified as a watershed decision support system (DSS), sponsored by EPRI. A DSS provides information and tools that help collaborative decision-making among interested parties (Chen *et al.* 2001b). WARMF was designed to assist in watershed management and TMDL development, and is publically available. WARMF intended users are technical and non-technical stakeholders making watershed decisions. WARMF is organized into five linked modules (Engineering, Data, Consensus, TMDL, and Knowledge) under one, GIS-based GUI. The Engineering module is the dynamic, simulation model that drives WARMF. The Data module provides time series input data (meteorological, point source) and calibration data. The Knowledge module is a utility to store important documents for the watershed. At the center of WARMF are the two watershed approach modules for Consensus building and TMDL calculations, which provide road maps for the step-by-step decision-making process (Weintraub *et al.* 2001a). The model can be used to run simulations for certain management goals and objectives within a watershed that enables the user to see and compare the outcomes of alternative management plans. Output from the model can be shown in GIS-based maps, graphs, and tables.

The algorithms of WARMF were derived from many well-established codes (Chen *et al.* 2001b). Algorithms for snow hydrology, groundwater hydrology, river hydrology, lake hydrodynamics, and mass balance for acid-base chemistry were based on the Integrated Lake-Watershed Acidification Study model (Chen *et al.* 1983). Algorithms for erosion, deposition, resuspension, and transport of sediment were adapted and modified from ANSWERS (Beasley



et al. 1980; Beasley & Huggins 1981). The pollutant accumulation and wash-off from urban areas was adapted from the Storm Water Management Model (Chen & Shubinski 1971; USEPA 1992). The sediment sorption-desorption of pesticides and phosphorus and the kinetics of nutrients and algal dynamics were adapted from WASP5 (Ambrose et al. 1991). A complete description of the WARMF formulations can be found in Chen et al. (1998).

### Engineering module

The Engineering module is a GIS-based watershed model that calculates daily runoff, subsurface flow, stream flow, and water quality of river segments and reservoirs. The model divides the watershed into various components, including sub-watersheds, stream segments, and lake layers. Figure 1 shows the network of sub-watersheds and rivers for the St. Esprit Watershed. Sub-watersheds are further divided into canopy and soil layers. Land surface is described by land use. In order to run water quality simulations, these components are connected into an integrated network allowing for the flow of pollutants between them.



Figure 1 | Network of land catchments and rivers for the St. Esprit Watershed.

A hydrologic model within WARMF simulates canopy interception, snow pack accumulation and melt, infiltration through soil layers, ET from soil, ex-filtration of groundwater to stream segments, and kinematic wave routing of stream flows. Figure 2 shows the conceptual model of hydrology for WARMF. A sub-watershed can have various land uses on the land surface. Below ground, the soils can have up to five layers (only two layers are shown in Figure 2). The groundwater table can rise or fall depending on the balance between vertical percolation from above and lateral outflow to the river segment (Chen et al. 2001b). The potential ET for each month is calculated as a function of latitude using Hargreaves' equation (Hargreaves 1974).

In general, the hydrologic simulation is performed as follows (Chen et al. 2005). The rate of infiltration into the soil is limited by the vertical hydraulic conductivity of the top soil layer. If the soil is frozen (which occurs at the St. Esprit Watershed) or the groundwater table rises to the ground level, the water from precipitation and snowmelt is backed up to the ground surface. The water retained on the ground surface fills surface depression storage. When surface depression storage is filled, the excess water flows to a river segment by sheet flow, which is calculated by Manning's equation. For each soil layer, percolation is calculated from the layer above and to the layer below. The soil layer has an allocation of evaporation according to the root distribution of plants. The model performs a flow balance in each time step (daily) to update soil moisture



Figure 2 | Definition sketch for the compartments of a watershed (Chen et al. 2001b).



by accounting for the evaporation and the difference in percolation to and from the soil layer. The hydraulic conductivity of soil layers is a function of soil moisture. The available void space in the soil is filled with the water, which percolates downward. When the percolation reaches the groundwater table, it raises the groundwater level. Groundwater can flow out to the river segment by lateral flow, which is calculated with Darcy's equation using horizontal hydraulic conductivity and slope. The unconfined aquifer is assumed to be watertight. Any known loss of groundwater to the deep confined aquifer must be specified as groundwater pumping for WARMF to extract water from the unconfined aquifer. Stream flow is routed by the kinematic wave method, and Manning's equation is used to calculate the outflow rate. Water depth is determined by performing a flow balance by accounting for inflow from the upstream river segment, inflow from land catchments on both sides of the river segment, outflow to the downstream river segment, and the change of storage. Such simulations track the flow paths of precipitation from land into different water bodies (Chen *et al.* 2005).

The Chemistry module performs various mass balance and chemical equilibrium calculations along each flow path (Weintraub *et al.* 2001b; Eisen-Hecht & Kramer 2002). A complete mass balance is performed, starting with atmospheric deposition and land application as boundary conditions. Pollutants are routed with water in throughfall, infiltration, soil adsorption, exfiltration, and overland flow. The sources of point and non-point loads are routed through the system with the mass so that the source of non-point loading can be tracked back to land use and location (Chen *et al.* 1998).

### Data and Knowledge modules

The Data module contains meteorology, air quality, and point source data used to run the model. It also contains observed flow and water quality data used for model calibration purposes. The data are accessed using a map-based interface and can be viewed and edited in both graphical and tabular format. Supplemental watershed data, documents, case studies, or reports of past modeling activities are stored in the Knowledge module for easy access by model users (Chen *et al.* 1998).

### Consensus and TMDL modules

The last two watershed approach modules are roadmaps providing guidance for stakeholders during the decision-making process. The Consensus module provides information in a series of steps for stakeholders to learn about the issues, formulate and evaluate alternatives, and negotiate a consensus. It provides a simple menu for scenario generation that allows stakeholders, without extensive WARMF knowledge, to simulate reductions of loads from point/non-point sources of pollution, atmospheric deposition, or diversion quantities by a percentage (Chen *et al.* 2001b). Users require knowledge of the models and interfaces to run more detailed scenarios (e.g., changes in land use distribution, changes in fertilizer application rates, etc.). Through the TMDL module, calculations are made for a series of control points from upstream to downstream within the watershed. Iterative sets of simulations can be performed to calculate various combinations of point and non-point loads that the water body can accept and meet water quality criteria for the designated uses at these control points (Chen *et al.* 1998).

### Site description

The WARMF model was applied to the St. Esprit Watershed, located approximately 50 km north of Montreal between 45°55'0" and 46°0'0" N, and 73°41'32" and 73°36'0" W (Figure 3) in south-western Quebec, Canada. It is a part of the 210 km<sup>2</sup> St. Esprit River basin, a tributary of the L'Assomption Watershed (4,220 km<sup>2</sup>). The land in the watershed is predominantly used for agriculture. The human population of the watershed is about 700; however, there are no villages or towns within the watershed.

The St. Esprit Watershed comprises a net drainage area of 24.3 km<sup>2</sup>. During the study period (1994–1996), approximately 64% of the total area was under crop production with the majority of land use under corn crop, followed by cereals, soybeans, vegetables, hay, and pastures. The remaining 36% of the area was occupied by forested, bare, and residential lands (Table 1). Over 50% of the agricultural land has subsurface drainage. The difference in elevation from the outlet to the highest point of the watershed is

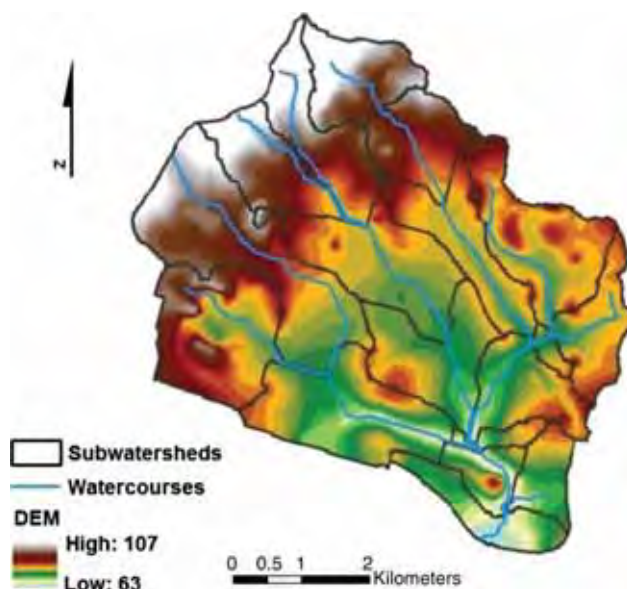


**Figure 3** | Location of St. Esprit Watershed.

**Table 1** | Land use distribution on St. Esprit Watershed

Land use	Area (m <sup>2</sup> )	Area (%)
Corn	6,272,063	26
Cereal	2,073,508	8.5
Soya	1,565,615	6.4
Vegetable	1,966,098	8.1
Hay	2,919,677	12
Forest	6,345,533	26
Pasture	703,183	2.9
Irrigation-pond	165,435	0.68
Residential	1,241,621	5.1
Unused	1,107,939	4.5
Total	24,360,673	100

44 m and the principal watercourse is 8.5 km long. Topography can be described as flat to rolling, with most of the cultivated land having slopes of less than 3%. The elevation data were obtained from GeoBase (GeoBase 2007), Canadian Digital Elevation Data; the watershed boundary was created using GIS tools and digital elevation map (DEM) (Figure 4). The streams map was taken from previous studies on St. Esprit Watershed and also created using DEM and GIS tools (Enright *et al.* 1995; Mousavizadeh *et al.* 1995; Sarangi *et al.* 2005a, b).



**Figure 4** | Elevation map of St. Esprit Watershed (DEM) (sub-watersheds are delineated using the streams network).

These maps, along with 1:63,360 soil maps (Lajoie 1965), 1:15,000 field-level aerial photography, and information provided by Enright *et al.* (1995) identified approximately 16 soil types, and 10 different land use categories (Figure 5). Soil textures in the watershed are variable; in general, the largest proportion of the watershed is occupied by coarse-textured

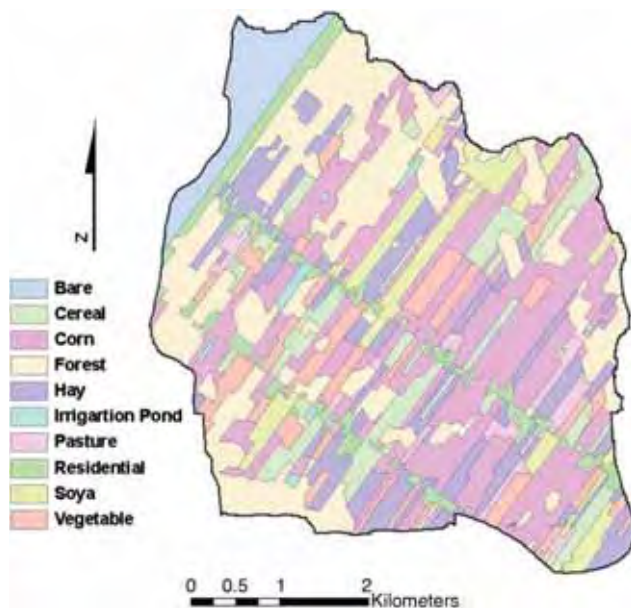


Figure 5 | Land use map of St. Esprit Watershed.

soils (sand and sandy loam 44%), followed by fine-textured soils (clay and clay loam 39%). The distribution of soil textural classes in the watershed is shown in Figure 6. The lower portion of the watershed is mostly composed of clays and clay loams, including the Ste. Rosaile and St. Laurent Series (Lapp *et al.* 1998). Most of the annual

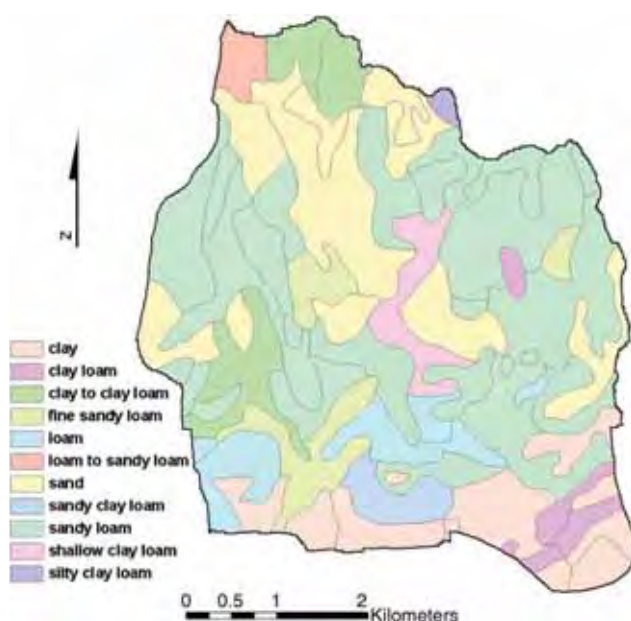


Figure 6 | Soil texture map of St. Esprit Watershed.

crop production takes place on the heavier soils. The upper regions of the watershed are composed of loamy and sandy soils. Natural drainage on these soils is poor and tile drainage systems have been installed to lower the water table to enhance crop production (Enright *et al.* 1995).

In the watershed, the period of frost varies from 122 to 138 days. Average annual precipitation varies between 860 and 1,050 mm, with approximately 20 to 25% as snow (Saranghi *et al.* 2005a, b). Average annual potential ET is between 400 and 560 mm. The mean temperature in the month of July, the warmest month of the year, varies between 18 and 21 °C (MAPAQ 1983).

### Instrumentation and monitoring

In the winter of 1993–1994, a stream gauging station was established at the watershed outlet, and a meteorological station was installed in the watershed (Figure 3). The equipment installed at the watershed outlet included a water level sensor (Druck 950 submersible pressure transducer) installed on the stream bed bottom, a UDG01 ultrasonic level sensor mounted over the outlet culvert, and a data logger (Campbell Scientific CR10) located in the gauging station building to record and store the data. A backup system that independently measures water level and flow velocity and sends these data to the primary data logger was also installed. Water samples were collected on a flow-weighted basis. An automated water sampler was installed at the gauging station. A sampler intake line was suspended over the control section to be monitored. An automated sampling strategy was based on the flow volume calculation; the automated sampler was programmed for activation at a variable but predetermined threshold value of accumulated flow. The collected samples consisted of the automated type and the in-stream grab samples collected on the weekly or bi-weekly site visits. Details of the water sampling protocol are given in Enright *et al.* (1995).

The meteorological station was equipped with sensors for air and soil temperature, solar radiation, wind speed and direction, snow accumulation, as well as a tipping bucket rain gauge and a Campbell Scientific data logger (Perrone & Madramootoo 1998). Table 2 shows monthly precipitation for 1994–1996 measured at the site and



**Table 2** | Monthly precipitation (mm)

	1994	1995	1996	30-year average
January	111.0	140.0	156.0	66.5
February	45.0	37.9	131.0	37.0
March	47.3	53.8	50.0	52.0
April	105.2	44.3	73.0	78.1
May	53.3	85.0	53.0	86.7
June	187.0	38.0	92.0	103.2
July	119.0	122.0	96.0	87.5
August	113.0	82.0	85.0	92.1
September	45.0	58.0	89.0	98.0
October	20.0	91.0	70.0	87.0
November	63.0	97.0	71.0	89.9
December	19.0	45.0	46.0	63.4
Total	927.8	893.1	1012.0	941.4

average 30-year monthly precipitation measured at the St. Jacques weather station, located about 6 km from the experimental site. The annual precipitations of 1994 and 1995 were 14 and 49 mm below the 30-year average (1997–2007), respectively; while, the annual precipitation in 1995 was 71 mm above the 30-year average range of monthly average precipitation for 30 years = 37 to 103.2 mm (Table 2).

Land use and land management information were collected on St. Esprit as part of an integrated watershed monitoring and management project (Enright *et al.* 1995; Papineau & Enright 1997). There were 25 farms, of which information from 18 of the farms was available. The participating producers account for approximately 67% of the agricultural land use of the watershed (Sarangi *et al.* 2005a, b).

### Model inputs

WARMF inputs include meteorology data (daily precipitation and min/max temperature, cloud cover, dew point temperature, air pressure, and wind speed), soil properties (field and saturated moistures of soil, vertical and horizontal hydraulic conductivities, and bulk density), land use data, DEM, sub-watershed boundaries, ground slope and aspect, and fertilizer application data. WARMF's water quality-related parameters comprise initial soil concentration of

ammonia, soil nitrification rate, litter fall rate, productivity of land uses, and air quality data.

The climatic data were obtained from the meteorological station installed in the watershed. The DEM (Figure 4) was developed using topographic data (contour lines and elevation point data). The natural drainage network was generated from the DEM of the watershed, using the Arc-Hydro tools of GIS. The sub-watershed boundary map (watershed delineation) was developed using DEM and by knowing where the watershed outlet is located. Using GIS tools, the watershed was discretized into 18 sub-watersheds ranging from several hundred m<sup>2</sup> to a few km<sup>2</sup> in size (Figure 1). The slope and aspect maps were developed using DEM in GIS. The county-level soil maps were digitized and a soil database (Figure 6) was developed and used to identify the different soil types across the watershed. Crop-related data were taken from land use maps imported into GIS (Figure 5). Fertilizer data were obtained from the GIS shape files developed for the watershed (Papineau & Enright 1997; Mousavizadeh 1998). WARMF overlays the polygons of land uses to the polygons of sub-watersheds to calculate the percentages of land use categories (e.g., urban, agriculture, deciduous forest, coniferous forest, open, etc.). The van Genuchten equation (Van Genuchten 1980) was used to determine the soil parameters (retention curve and hydraulic conductivity). The van Genuchten parameters were obtained from previous research performed in Quebec (Mousavizadeh 1992, 1998; Perrone 1997; Dayyani *et al.* 2009, 2010) and also using the Rosetta model (Schaap *et al.* 2001). Air and rain chemistry data were obtained from the National Atmospheric Deposition Program (NADP) website (NADP 2009). The closest station to the watershed was HWF187, located approximately 220 km south of the watershed. Other parameters were set to the default values recommended in the model manual.

### Methods of evaluation

Three years of data (1994–1996) were available for the study area. This data set was divided into two parts: 1994–1995 for calibration and 1996 for validation. Statistical tools such as average deviation (AD), root mean square error (RMSE), coefficient of determination ( $R^2$ ), and Nash–Sutcliffe ( $E$ ) coefficient (Nash & Sutcliffe 1970), were used to evaluate

model performance for both calibration and validation periods. The coefficient of determination is a measure of accuracy or the degree to which the measured and predicted values agree. The average deviation is used to determine whether the model made over- or underpredictions. The RMSE measures the difference between predicted and observed values. It is sensitive to the extreme values and deals with both systematic and random errors. The Nash and Sutcliffe coefficient measures the goodness-of-fit between observed and simulated values. A value of 1 represents a perfect match, while a value of zero (0) shows a prediction no better than using the mean of the data. The negative efficiency indicates that the prediction is worse than simply taking the mean of the measured values.

Moriasi *et al.* (2007) reviewed performance statistics reported by different researchers for many watershed models and summarized model evaluation criteria. The model performance ratings vary from 'unsatisfactory' to 'very good' for Nash–Sutcliffe efficiency ( $E$ ) (very good: 0.75–1; good: 0.65–0.75; satisfactory: 0.5–0.65; unsatisfactory:  $\leq 0.5$ ), RMSE observations, standard deviation ratio (RSR), and percent bias (PBIAS). Coefficient of determination ( $R^2$ ) and  $E$  are the most widely used statistics for hydrological modeling performance evaluation (Gassman *et al.* 2007) and typically  $R^2$  values greater than 0.5 are considered acceptable (Santhi *et al.* 2001).

The model performance was first qualitatively assessed with graphical displays of the results and then statistical measures were used for quantitative evaluation. The graphical comparison is made easily using the scenario manager of the WARMF model. For example, Scenario 1 may be used to represent a set of numerical values of model coefficients used in the simulation. Scenario 2 may be used to represent a second set of modified model coefficients used in the simulation. Once the simulations are performed, WARMF can plot the observed data as well as the model predictions for both scenarios on the same graph. By visual inspection, it is relatively easy to see whether the changes to model coefficients improve the match.

### Model calibration and validation

Model calibration is the procedure in which model parameters are adjusted to improve the match between the

simulated and observed values. Only a few parameters are adjusted for each calibration. In this study, the model was calibrated by adjusting the evaporation coefficients (magnitude, skewness), snow/ice-related parameters (melting rates, snow formation, and melting temperature), field and saturated soil moisture, vertical/horizontal hydraulic conductivities, detention storage, initial soil concentration of nitrate, surface roughness (Manning's  $n$ ), litter fall rate, and soil nitrification and denitrification rates. These parameters were selected for calibration based on the literature (Weintraub *et al.* 2001b; Geza & McCray 2007; Herr 2008, personal communication). The initial values of the calibration parameters were taken from studies conducted at the St. Esprit Watershed (Mousavizadeh *et al.* 1995; Perrone 1997; Mousavizadeh 1998; Sarangi *et al.* 2005a, b) or studies performed in regions with the same soil and climatic conditions (Dayyani *et al.* 2009, 2010).

Measured flow, nitrate–nitrogen ( $\text{NO}_3^-$ -N), and climatic data are available from 1994 to 1996 for the gauging and weather stations installed in the watershed. Model initialization is important as WARMF assumes zero initial snowpack so it is recommended to start simulation before January 1. Thus, the calibration started on May 1, 1993 in order to initialize the model; the first 8 months of simulation (before January 1, 1994) were not considered in the evaluations of the calibration process. The meteorological data for these extra 8 months were obtained from St. Jacques weather station, since it was not recorded at the weather station installed in the watershed. Ideally, one may want to collect field data in a variety of wet and dry years. However, the field investigators have no control over the natural meteorologic and hydrologic variability. Model calibration/validation must be performed for the period that the field data were collected.

## RESULTS AND DISCUSSION

### Hydrologic simulations

WARMF calculates daily flow and pollutant concentrations for all river segments. As in this study the data were measured only at the outlet of the watershed, the simulated



flow and  $\text{NO}_3\text{-N}$  concentrations were compared with the observed values at the outlet of the watershed.

Figure 7 compares simulated and observed daily hydrographs at the outlet of the watershed for the entire simulation period of 1994–1996. Figure 7(a) is for the calibration period of 1994–1995, in which the simulated and observed daily values showed an  $R^2$  value of 0.53 and E value of 0.45 (Table 3); Figure 7(b) shows results for the validation period with the  $R^2$  of 0.58 and E of 0.5. The model seems to have performed reasonably well at simulating the timing of runoff events. WARMF underestimated several peaks during the snowmelt for both calibration and validation periods. The statistical indices calculated from the predicted and observed daily drainage outflows are given in Table 3. The positive mean deviation during the calibration and validation period between observed and simulated daily values indicates that the model slightly underpredicted flow. Mean deviation values of daily drain flow, 0.50 and 0.29 mm, indicate that daily underestimations were small.

In certain cases during the winter/spring period, the model simulated greater flow than was observed which could be due to the fact that the model does not consider frozen water on the ground surface, eventually simulating greater runoff. For most of the events during the

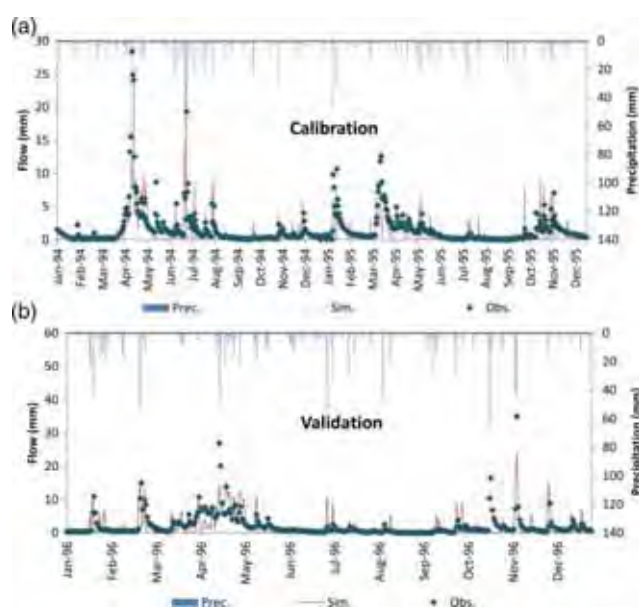


Figure 7 | Daily hydrograph and hydrographs for calibration (a) and validation (b) periods.

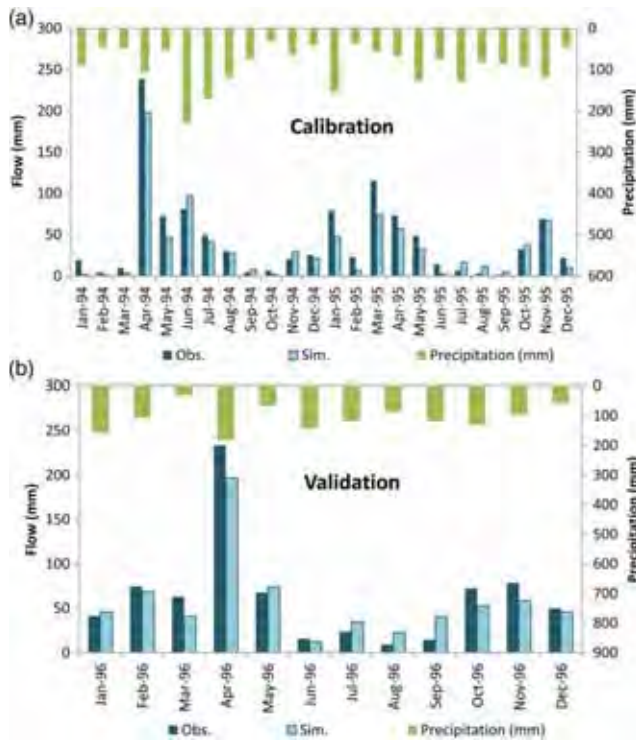
Table 3 | Model performance during calibration and validation

Statistical parameter	Calibration (1994–1995)	Validation (1996)
Daily drain flow		
AD (mm)	0.50	0.29
RMSE (mm)	2.37	3.26
$R^2$	0.53	0.58
$E$	0.45	0.50
Monthly drain flow		
AD (mm)	7.27	4.16
RMSE (mm)	17.06	17.62
$R^2$	0.92	0.94
$E$	0.89	0.91
Monthly nitrogen losses		
AD ( $\text{kg}\cdot\text{ha}^{-1}$ )	0.18	0.41
RMSE ( $\text{kg}\cdot\text{ha}^{-1}$ )	0.48	0.86
$R^2$	0.89	0.84
$E$	0.86	0.75

winter/spring period, simulated and observed timing of runoff peaks was found to differ by 1 day. This difference is thought to arise due to the quantity of accumulated snow on the soil surface resulting in differences in the roughness of bare and snow-covered land surfaces. The model considers the same surface roughness coefficient irrespective of the day of the year or the soil surface conditions. In reality, flow velocity is higher on snow-covered surfaces due to lower friction (in a cold climate the surface of snow-pack is usually frozen); thus, the time for the flow to reach the watershed outlet is faster.

Figure 8 presents WARMF's simulated monthly results and observed data. The WARMF calibration results showed a high correlation between simulated and observed monthly flows ( $R^2 = 0.92$  and  $E = 0.89$ ), although the model seemed to slightly overestimate flow during a few months ( $\text{AD} > 0$ ; Table 3). During the validation period, the  $R^2$  and  $E$  were 0.94 and 0.91, respectively, showing good correspondence between the simulated and observed monthly values. The modeling errors (RMSE) during calibration and validation years were low, 17.1 and 17.6 mm, further indicating that the model performed well in predicting monthly flow at the outlet of the watershed.

As mentioned earlier, overall the model seems to underestimate the flow during the winter and snowmelt period.



**Figure 8** | Monthly hydrograph and hydrographs for calibration (a) and validation (b) periods.

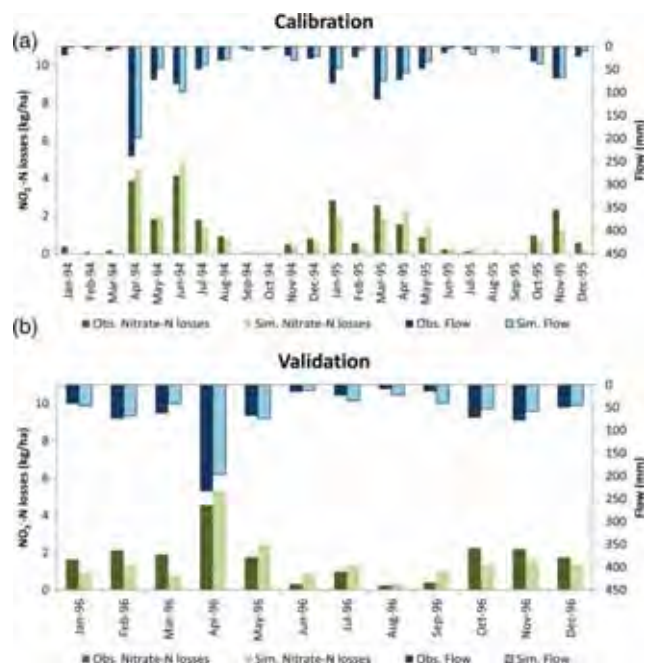
Other researchers in the Quebec region have reported that there is considerable amount of flow coming out of the tile drainage system during the winter and spring (Gollamudi 2006; Dayyani *et al.* 2009, 2010). As WARMF does not take into account the tile drainage system, and almost half of the St. Esprit Watershed is subsurface drained, it does not simulate the base flow properly especially during the winter and snowmelt period, leading to underestimating the total flow during those months.

The model validation results are slightly better than the model calibration results (Table 2; higher  $R^2$  and  $E$  values). This is contrary to what is normally observed in model simulations; calibration results are usually better than validation results. While it may indicate that the model has not been overly calibrated, it is also likely that the WARMF model may be simulating a wetter year better. From Table 2, one can see that not only 1996, the validation year, was wetter than both 1994 and 1995, the two calibration years, it was also wetter as compared to the 30-year average precipitation for the watershed. However, this observation needs to be checked further in future modeling studies with this model.

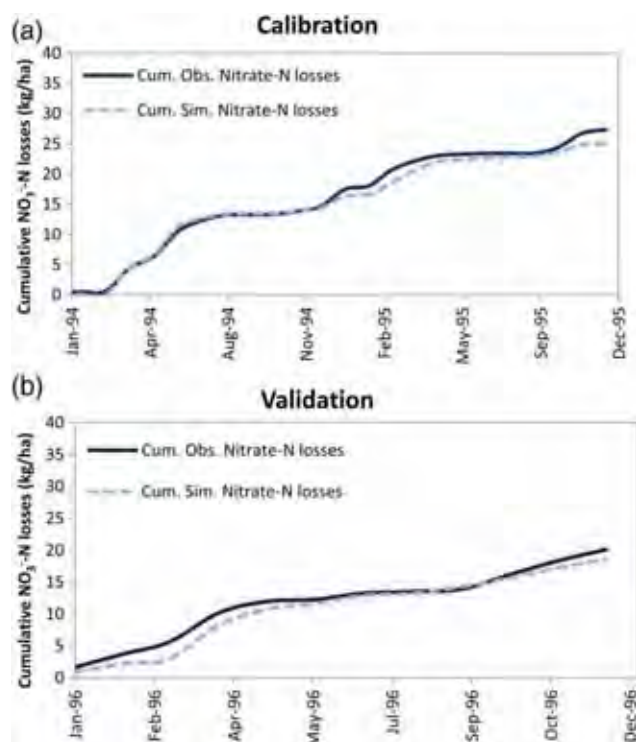
## Nitrogen simulations

Once the hydrology component of the model was calibrated and validated, nitrogen simulations were conducted to calibrate the nitrogen component of the model. Water-quality data collected during the flow calibration period, 1994–1995, were used to guide the nitrogen calibration. Using parameters fine-tuned in the calibration process, the model was validated using 1996 data.

The fertilizer application rate and timing of data were taken from the GIS shape files and St. Esprit Watershed reports (Papineau & Enright 1997; Mousavizadeh 1998). The adjusted parameters used during calibration were soil nitrification and denitrification rates, litter fall rate, and initial  $\text{NO}_3^-$ -N concentrations in the soil. Initial runs of the model to simulate  $\text{NO}_3^-$ -N concentrations were conducted using the nitrogen input values, taken from Dayyani *et al.* (2010), Elmi *et al.* (2000), and Lapp (1996). Figure 9 compares the simulated and observed monthly  $\text{NO}_3^-$ -N losses at the outlet of St. Esprit Watershed. The movement of  $\text{NO}_3^-$ -N has been closely associated with the movement of water in agricultural soils (Armstrong & Burt 1993). In this study, the  $\text{NO}_3^-$ -N losses were also dependent



**Figure 9** | Simulated versus observed monthly  $\text{NO}_3^-$ -N losses at the outlet for calibration (a) and validation (b) periods.



**Figure 10** | Simulated versus observed cumulative  $\text{NO}_3\text{-N}$  losses at the watershed outlet for calibration (a) and validation (b) periods.

on outflow rates (Figure 9). The comparison between the observed and simulated monthly  $\text{NO}_3\text{-N}$  and cumulative losses (Figures 9 and 10) shows that the model performed quite well with the  $R^2$  values of 0.89 and 0.84 and E values of 0.86 and 0.75 for calibration (Figure 9(a)) and validation (Figure 9(b)) periods, respectively (Table 3).

About 67% of the overall  $\text{NO}_3\text{-N}$  losses was recorded in April, May, and June 1994; this is correlated to the intensive flow rates (72% of the annual flow) measured during this period (Figure 9(a)). In 1994, heavy rainfall events occurred soon after the second surface fertilizer application (early June). June had 187 mm of rainfall, 1.8 times the 30-year average rainfall (Table 2). Although the flow was being underestimated over the snowmelt period in all three years, the  $\text{NO}_3\text{-N}$  losses were overestimated (Figures 9(a) and 9(b)), which might be because of the timing of first and second fertilizer application, in May and June. Overall, the comparison between the observed and simulated monthly and cumulative  $\text{NO}_3\text{-N}$  losses (Figure 10) shows that the model performed quite well, although tended to slightly underestimate the total annual losses by 9%

and 8% during calibration (Figure 10(a)) and validation (Figure 10(b)) periods, respectively. This could be due to the flow underestimation by the model.

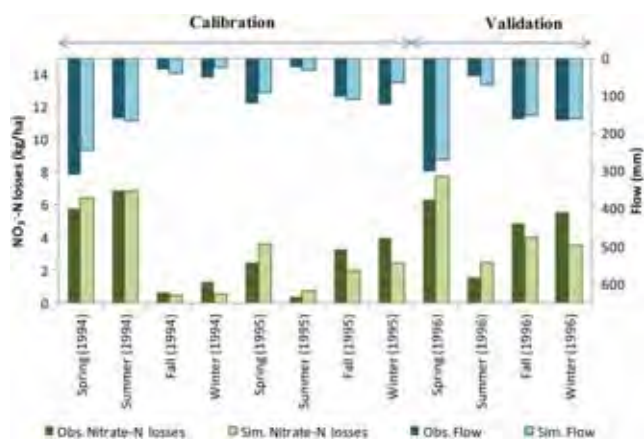
The statistical comparison (Table 3) showed good agreement between simulated and measured monthly  $\text{NO}_3\text{-N}$  losses during the study period. The overall positive mean deviation (AD) value (0.18 and  $0.41 \text{ kg ha}^{-1}$ ) indicates that the monthly N losses are slightly underestimated (Table 3). The coefficient of determination was more than 0.84, indicating a good correspondence between simulated and observed  $\text{NO}_3\text{-N}$  losses. Overall, the model performance was slightly better during the calibration period, with the modeling efficiency (E) being 0.86 as compared to 0.75 for validation.

The comparison between the observed and simulated annual nitrate–nitrogen losses (Table 4) shows that the model performed quite well, although tended to slightly underpredict cumulative losses, which could be due to underestimation of flow.

The WARMF performed reasonably well in simulating the seasonal variation of flow (Figure 11), although the

**Table 4** | Observed and simulated cumulative flow and  $\text{NO}_3\text{-N}$  losses for calibration and validation periods

	Observed		Simulated	
	Flow (mm)	N ( $\text{kg ha}^{-1}$ )	Flow (mm)	N ( $\text{kg ha}^{-1}$ )
1994	563.8	14.6	478.3	14.4
1995	488.5	12.7	371.1	10.7
1996	742.4	20.1	692.5	18.6



**Figure 11** | Simulated versus observed seasonal flow and  $\text{NO}_3\text{-N}$  losses at the watershed outlet for calibration (1994–1995) and validation (1996) periods.



spring flow is underestimated in all years. The dynamics of nutrient transport are different in the three seasons which experienced high flows (winter, spring, and fall), due to the changing hydrologic conditions among these seasons. WARMF was able to reproduce the conditions of all seasons satisfactorily for both the calibration and validation years, although the spring nitrate–nitrogen loads are overestimated (Figure 11).

Mousavizadeh (1998) applied ANSWERS2000 to the St. Esprit Watershed. The model seemed to underestimate the flow; total simulated cumulative runoff values were 66.6, 54.9, and 71.7% of measured cumulative runoff values, for 1994, 1995, and 1996, respectively. Romero (2006) evaluated the performance of the SLURP hydrological model at St. Esprit Watershed and reported  $R^2$  values of 0.522 and 0.66 for calibration and validation periods, respectively. Fernandez *et al.* (2002) evaluated the performance of the WHATGIS watershed model in simulating the monthly flow and nitrate-N loads. They reported Nash–Sutcliffe coefficients of 0.85 and 0.83 for flow and nitrate-nitrogen monthly simulations, respectively. Geza & McCray (2007) applied the WARMF model to a watershed in Colorado and reported  $E = 0.58$  and  $R^2 = 0.68$  for simulation of monthly stream flow. SWAT model has been applied to different watersheds around the world; the range of reported  $R^2$  and  $E$  values are for daily flow  $R^2 = 0.37–0.78$  (Spruill *et al.* 2000; Wang & Melesse 2005, 2008) and  $E = -0.04$  to 0.19 (Spruill *et al.* 2000); for monthly flow  $R^2 = 0.67–0.92$  (Santhi *et al.* 2001; Jha *et al.* 2007; Wang *et al.* 2008) and  $E = 0.54–0.94$  (Peterson & Hamlett 1998; Saleh *et al.* 2000; Di Luzio *et al.* 2002; Jha *et al.* 2007); for monthly simulation of nitrate-nitrogen  $R^2 = 0.72–0.89$  (Santhi *et al.* 2001; Jha *et al.* 2007) and  $E = 0.27–0.73$  (Saleh *et al.* 2000; Di Luzio *et al.* 2002; Jha *et al.* 2007). Comparing the results of this study (Table 3) with the above-mentioned applications of different watershed models indicates that the WARMF model performed well in simulating flow and N loads at the outlet of the watershed.

## CONCLUSIONS

The WARMF model was applied to a watershed in south-western Quebec, Canada, and adequately predicted the daily flow and nitrate-nitrogen losses at the outlet of the

watershed. The simulations were performed for a 3-year period from 1994 to 1996. The model simulated the pattern of monthly flow with a good degree of accuracy: the  $R^2$  for calibration was 0.92, and that for validation was 0.94. The corresponding coefficients of efficiency ( $E$ ) were 0.89 and 0.91. The model also performed well in simulating the flow during the snowmelt period with the error of 26%, 33%, and 11% in 1994, 1995, and 1996, respectively. The model also performed well in simulating nitrate loads at the outlet of watershed (Nash and Sutcliffe coefficient  $\geq 0.75$ ). We are satisfied with the performance of the WARMF model in a small agricultural watershed in a cold climate.

## REFERENCES

- Ambrose, R. B., Wool, T. A., Martin, J. L., Connolly, J. P. & Schanz, R. W. 1991 *WASP5x, a Hydrodynamic and Water Quality Model - Model Theory, User's Manual, and Programmer's Guide*. Environmental Research Laboratory. Environmental Protection Agency. Athens, GA.
- Armstrong, A. C. & Burt, T. P. 1993 Nitrate losses from agricultural land. In: *Nitrate: Processes, Patterns and Management* (T. P. Burt, A. L. Heathwaite & S. T. Trudgill, eds). Wiley, Chichester, UK, pp. 239–267.
- Beasley, D. B. & Huggins, L. F. 1981 Answers – user manual. In: EPA-905/9-82-001, US Environmental Protection Agency, Region V, Chicago, IL, p. 54.
- Beasley, D. B., Huggins, L. F. & Monke, E. J. 1980 ANSWERS – a model for watershed planning. *Trans. ASAE* **23** (4), 938–944.
- Chen, C. W. & Shubinski, R. P. 1971 Computer simulation of urban storm water runoff. *J. Hydraul. Div. ASCE* **97**, 289–301.
- Chen, C. W., Gherini, S. A., Judson, R. J. M. & Dean, J. D. 1983 Integrated Lake-Watershed Acidification Study, Volume 1: Model Principles and Application Procedures. Final Report, Electric Power Research Institute, Palo Alto, CA, EA-3221, Volume 1.
- Chen, C. W., Herr, J. W., Goldstein, R. A., Ice, G. & Cundy, T. 2005 Retrospective comparison of watershed analysis risk management framework and Hydrologic Simulation Program–Fortran applications to Mica Creek watershed. *J. Environ. Eng. ASCE* **131**, 1277–1284.
- Chen, C. W., Herr, J. & Ziemelis, L. 1998 Watershed Analysis Risk Management Framework – A Decision Support System for Watershed Approach and TMDL Calculation. Documentation Report TR110809. Electric Power Research Institute, Palo Alto, CA.
- Chen, C., Weintraub, L., Goldstein, R., Siegrist, R. & Kirkland, S. 2001a Framework to account for onsite wastewater systems in calculating total maximum daily loads. In: *On-Site*

- Wastewater Treatment: Proceedings of the Ninth National Symposium on Individual and Small Community Sewage Systems*. ASAE Publication, St Joseph, MI, pp. 524–532.
- Chen, C. W., Herr, J. & Weintraub, L. 2001b *Watershed Analysis Risk Management Framework (WARMF): Update One – A Decision Support System for Watershed Analysis and Total Maximum Daily Load Calculation, Allocation and Implementation*. Publication No. 1005181, Electric Power Research Institute (EPRI), Palo Alto, CA.
- Dayyani, S., Madramootoo, C., Enright, P., Simard, G., Gullamudi, A., Prasher, S. & Madani, A. 2009 [Field evaluation of DRAINMOD 5.1 under a cold climate: simulation of daily midspan water table depths and drain outflows 1](#). *JAWRA J. Am. Water Resour. Assoc.* **45** (3), 779–792.
- Dayyani, S., Prasher, S. O., Madramootoo, C. A. & Madani, A. 2010 Modeling Water Table Depths, Drain Outflow, and Nitrogen Losses in Cold Climate Using DRAINMOD 5.1. *Trans. ASABE, Am. Soc. Agric. Eng.* **53** (2), 385–395.
- Di Luzio, M., Srinivasan, R. & Arnold, J. 2002 [Integration of watershed tools and SWAT model into BASINS](#). *J. Am. Water Resour. Assoc.* **38** (4), 1127–1142.
- Eisen-Hecht, J. & Kramer, R. 2002 [A cost-benefit analysis of water quality protection in the Catawba Basin](#). *J. Am. Water Resour. Assoc.* **38** (2), 453–465.
- Elmi, A. A., Madramootoo, C. A. & Hamel, C. 2000 [Influence of water table and nitrogen management on residual soil NO<sub>3</sub> and denitrification rate under corn production in sandy loam soil in Quebec](#). *J. Agric. Ecosyst. Environ.* **79**, 187–197.
- Enright, P. & Madramootoo, C. A. 2004 Phosphorus losses in surface runoff and subsurface drainage waters on two agricultural fields in Quebec. In: *Drainage VIII – Proceedings of the Eighth International Drainage Symposium* (R. A. Cooke, ed.). Published by ASAE – St. Joseph, MI, USA, pp. 160–170.
- Enright, P., Papineau, F., Madramootoo, C. & Leger, E. 1995 The impacts of agricultural production on water quality in two small watersheds. CSAE Paper. Paper No. 95-101. CSAE/SCGR, Mansonville, QC.
- EPRI 2000 *Peer Review of the Watershed Analysis Risk Management Framework (WARMF): An Evaluation of WARMF for TMDL Applications by Independent Experts Using USEPA Guidelines*. TR1000252. Electric Power Research Institute, Palo Alto, CA.
- Fernandez, G., Chescheir, G., Skaggs, R. & Amatya, D. 2002 WATGIS: a GIS-based lumped parameter water quality model. *Trans. ASAE* **45** (3), 593–600.
- Gassman, P. W., Reyes, R. R., Green, C. H. & Arnold, J. G. 2007 The soil and water assessment tool: historical development, applications, and future research directions. *Trans. ASABE* **50** (4), 1211–1250.
- Geza, M. & McCray, J. E. 2007 Modeling the effect of population growth on stream nutrient concentration in Turkey Creek Watershed using WARMF model. In: *On-site Wastewater Treatment: Proceedings of the 11th National Symposium on Individual and Small Community Sewage Systems*, Amer. Soc. Agricul. Biological Engr., St. Joseph, MI.
- GeoBase 2007 from. <http://www.geobase.ca/> (accessed 10 May 2007).
- Gollamudi, A. 2006 Hydrological and Water Quality Modeling of Agricultural Fields in Quebec. MSc Thesis, McGill University, Montreal, Canada.
- Hargreaves, G. H. 1974 Estimation of potential and crop evapotranspiration. *Trans. ASAE* **17** (4), 701–704.
- Herr, J., Chen, C., Goldstein, R. & Brogdon, J. 2002 A tool for sediment TMDL development on Oostanaula Creek. In: *Proceedings of the Total Maximum Daily Load (TMDL) Environmental Regulations*. ASAE Publication Number 701P0102, pp. 111–116.
- Jha, M., Gassman, P., Arnold, J. & USDA, A. 2007 Water quality modeling for the raccoon river watershed using SWAT. CARD Working Paper 06-WP 428. Available at: [http://www2.econ.iastate.edu/research/webpapers/paper\\_12656.pdf](http://www2.econ.iastate.edu/research/webpapers/paper_12656.pdf).
- Keller, A., Zheng, Y. & Robinson, T. 2004 Determining critical water quality conditions for inorganic nitrogen in dry, semi-urbanized watersheds. *J. Am. Water Resour. Assoc.* **40** (3), 721–735.
- Lajoie, P. G. 1965 *Étude pédologique des comtés de L'Assomption et de Montcalm (Québec)*. Ottawa, ON, Canada Department of Agriculture. (Soil survey of the counties of L'Assomption and Montcalm.) Available at: [http://www.irda.qc.ca/\\_ftbFiles/Etude\\_pedo/Etude\\_pedo\\_32.pdf](http://www.irda.qc.ca/_ftbFiles/Etude_pedo/Etude_pedo_32.pdf).
- Lapp, P. 1996 The Hydrology and Water Quality of an Intensive Agricultural Watershed in Quebec. MSc Thesis, Department of Agriculture and Biosystems Engineering, McGill University, Montreal, Canada.
- Lapp, P., Madramootoo, C., Enright, P. & Perrone, F. P. J. 1998 [Water quality of an intensive agricultural watershed in Quebec](#). *J. Am. Water Resour. Assoc.* **34** (2), 427–437.
- MAPAQ, Q. 1983 *Agricultural Characteristic of the Montcalm County*. Regional Directorates of L'Assomption, 76 pp.
- Moriassi, D. N., Arnold, J. G., Van Liew, M. W., Binger, R. L., Harmel, R. D. & Veith, T. L. 2007 Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* **50** (3), 885–900.
- Mousavizadeh, M. H. 1992 Soil Physical Properties Measurement, Report. Ste. Anne-de-Bellevue, QC, Dept of Agricultural Engineering, McGill University, Macdonald Campus.
- Mousavizadeh, M. H. 1998 Integration of a Geographic Information System and a Continuous Nonpoint Source Pollution Model to Evaluate the Hydrologic Response of an Agricultural Watershed. PhD Thesis, Bioresource Engineering, McGill University, Montreal, Canada.
- Mousavizadeh, M. H., Papineau, F., Enright, P. & Madramootoo, C. A. 1995 Application of GIS and water quality models to watershed management. CSAE Paper.
- NADP 2009 *National Atmospheric Deposition Program (NADP) Website*. Available at: <http://nadp.sws.uiuc.edu/NADP/> (accessed 3 November 2008).
- Nash, J. E. & Sutcliffe, J. V. 1970 [River flow forecasting through conceptual models: Part I – A discussion of principles](#). *J. Hydrol.* **10**, 282–290.



- Papineau, F. & Enright, P. 1997 Gestion de l'Eau dans le Bassin Versant de la Partie Supérieure du Ruisseau St-Esprit, Project. Caractérisation de la problématique environnementale. Projet 61-13008. Plan Vert. Université McGill, Montréal, Québec.
- Perrone, J. T. 1997 Hydrologic Modeling of an Agricultural Watershed in Quebec Using AGNPS. MSc Thesis, Bioresource Engineering, McGill University, Montreal, Canada.
- Perrone, J. & Madramootoo, C. A. 1998 Improved curve number selection for runoff prediction. *Can. J. Civil Eng.* **25**, 728–734.
- Peterson, J. R. & Hamlett, J. M. 1998 Hydrologic calibration of the SWAT model in a watershed containing fragipan soils. *J. Am. Water Resour. Assoc.* **34** (3), 531–544.
- Rambow, B., Tufford, D. & Murray, J. 2008 Evaluating water quality trading for phosphorus in the lower catawba river basin. In: *Proceedings of the 2008 South Carolina Water Resources Conference*, North Charleston, SC 29208.
- Romero, D. R. 2006 Hydrologic Modeling on the Saint Esprit Watershed. MSc Thesis, McGill University, Montreal, Canada.
- Saleh, A., Arnold, J., Gassman, P., Hauck, L., Rosenthal, W., Williams, J. & McFarland, A. 2000 Application of SWAT for the upper North Bosque River watershed. *Trans. ASAE* **43** (5), 1077–1087.
- Santhi, C., Arnold, J., Williams, J., Dugas, W., Srinivasan, R. & Hauck, L. 2001 Validation of the SWAT model on a large river basin with point and nonpoint sources. *J. Am. Water Resour. Assoc.* **37** (5), 1169–1188.
- Sarangi, A., Madramootoo, C. A., Enright, P. & Chandrasekharan, H. 2005a Prediction of spatial variability of phosphorous over the St-Esprit Watershed. *Water Air Soil Pollut.* **168** (1), 267–288.
- Sarangi, A., Madramootoo, C., Enright, P., Prasher, S. & Patel, R. 2005b Performance evaluation of ANN and geomorphology-based models for runoff and sediment yield prediction for a Canadian watershed. *Curr. Sci.* **89** (12), 2022–2034.
- Schaap, M. G., Leij, F. J. & van Genuchten, M. T. 2001 Rosetta: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *J. Hydrol.* **251** (3–4), 163–176.
- Spruill, C., Workman, S. & Taraba, J. 2000 Simulation of daily and monthly stream discharge from small watersheds using the SWAT model. *Trans. ASAE* **43** (6), 1431–1439.
- USEPA (US Environmental Protection Agency) 1992 SWMM: Storm Water Management Model, Version 4.2. Center of Exposure Assessment Modeling. Environmental Protection Agency, Athens, GA.
- USEPA (US Environmental Protection Agency) 2009 Watershed Analysis Risk Management Framework (WARMF). Available at: <http://www.epa.gov/athens/wwqtsc/html/warmf.html> (accessed 12 November 2009).
- Van Genuchten, M. Th. 1980 A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J.* **44**, 892–898.
- Wang, X. & Melesse, A. 2005 Evaluation of the SWAT model's snowmelt hydrology in a northwestern Minnesota watershed. *Trans. ASAE* **48** (4), 1359–1376.
- Wang, X., Shang, S., Yang, W. & Melesse, A. 2008 Simulation of an agricultural watershed using an improved curve number method in SWAT. *Trans. ASAE* **51** (4), 1323–1339.
- Weintraub, L., Chen, C., Goldstein, R. & Siegrist, R. 2004 WARMF: a Watershed modeling tool for onsite wastewater systems. In: *Conference Proceedings of the On-Site Wastewater Treatment*, ASAE Publication Number 701P0104, St. Joseph, MI, pp. 636–646.
- Weintraub, L. H. Z., Chen, C. W. & Herr, J. 2001a Demonstration of WARMF: a decision support tool for TMDL development. In: *Proceedings of the Water Environment Federation (WEF) TMDL Science Issues Conference*, March 4–7, St. Louis, MO.
- Weintraub, L., Olmsted, L., Chen, C., Goldstein, R., Vaughan, G., Johnson, S., Ziegler, T., Foris, B., Brown, A. & Besler, D. 2001b Decision support system for Catawba river basin. In: *Proceeding of Integrated Decision-making for Watershed Management Symposium: Processes and Tools*, Virginia Water Resources Research Center, Blacksburg, VWRRC P7-2001, pp. 890–905.

# DNA extraction method selection for agricultural soil using TOPSIS multiple criteria decision-making model

Sepideh Pakpour<sup>1</sup>, Snizhana V. Olishevskaya<sup>2,3</sup>, Shiv O. Prasher<sup>2</sup>, Abbas S. Milani<sup>4</sup>,  
Martin R. Chénier<sup>3,5\*</sup>

<sup>1</sup>Department of Biology, University of British Columbia, Kelowna, Canada

<sup>2</sup>Department of Bioresource Engineering, McGill University, Montreal, Canada

<sup>3</sup>Department of Animal Science, McGill University, Montreal, Canada

<sup>4</sup>School of Engineering, University of British Columbia, Kelowna, Canada

<sup>5</sup>Department of Food Science and Agricultural Chemistry, McGill University, Montreal, Canada

Email: \*[martin.chenier@mcgill.ca](mailto:martin.chenier@mcgill.ca)

Received 20 June 2013; revised 22 July 2013; accepted 15 August 2013

Copyright © 2013 Sepideh Pakpour *et al.* This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## ABSTRACT

There is an increased interest in the extraction of nucleic acids from various environmental samples since culture-independent molecular techniques contribute to deepen and broaden the understanding of a greater portion of uncultivable microorganisms. Due to difficulties to select the optimum DNA extraction method in view of downstream molecular analyses, this article presents a straightforward mathematical framework for comparing some of the most commonly used methods. Four commercial DNA extraction kits and two physical-chemical methods (bead-beating and freeze-thaw) were compared for the extraction of DNA under several quantitative DNA analysis criteria: yield of extraction, purity of extracted DNA ( $A_{260/280}$  and  $A_{260/230}$  ratios), degradation degree of DNA, easiness of PCR amplification, duration of extraction, and cost per extraction. From a practical point of view, it is unlikely that a single DNA extraction strategy can be optimum for all selected criteria. Hence, a systematic Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) was employed to compare the methods. The PowerSoil<sup>®</sup> DNA Isolation Kit was systematically defined as the best performing method for extracting DNA from soil samples. More specifically, for soil:manure and soil:manure:biochar mixtures, the PowerSoil<sup>®</sup> DNA Isolation Kit method performed best, while for neat soil samples its alternative version gained the first rank.

**Keywords:** DNA Extraction; Agricultural Soil; Biochar; Poultry Manure; Multiple Criteria Decision-Making;

\*Corresponding author.

Technique for Order Preference by Similarity to Ideal Solution

## 1. INTRODUCTION

Soil is a unique ecosystem containing many different niches and creating favourable conditions for the development of different groups of microorganisms [1,2]. Since less than 1% of soil microorganisms can be grown in laboratory conditions using culture media [3,4] and the vast majority are not cultivable [5,6], a significant number of studies dealing with microbial diversity utilize molecular tools such as competitive PCR, real-time PCR, denaturing gradient gel electrophoresis (DGGE) and large-scale parallel-pyrosequencing based on the extraction of environmental nucleic acids [7-11].

Numerous procedures exist for the isolation and purification of DNA from soil [1,9,10,12-14]. Studies suggest that the selection of an appropriate extraction and purification procedure depends on the physical and chemical characteristics of the soil matrix, such as organic matter, clay content and pH value, as well as on different amendments like biochar and poultry manure used in agriculture for improvement of soil fertility [15-19]. For example, environmental samples such as soil and sediments often contain high levels of organic matter, especially humic acids and phenolic compounds, as well as heavy metals which can inhibit the activity of the *Taq* DNA polymerase in PCR [20-22] and reduce the specificity of DNA hybridization analysis [10,23,24].

Indirect and direct approaches have been developed for extracting nucleic acids from soil samples. Indirect extraction of DNA from soil samples is based on the following steps: dispersion of soil particles, separation of

the cells from soil particles by centrifugation according to sedimentation velocities, buoyant density or both, lysis of extracted cells and finally DNA purification [1,25,26]. The most commonly used methods for direct extraction of DNA are based on physical-chemical membrane disruption techniques (bead-beating and freeze-thaw) that may allow greater yields of microbial DNA recovery [9,12,14,15,20,27,28]. In spite of many advantages of direct methods, they generally show lower DNA purity with a high degree of DNA shearing that can negatively affect PCR efficiency or specificity [1,20,27]. Due to the fact that some of the steps of DNA extraction using these direct methods might make DNA isolation expensive or impractical for processing the large number of samples usually required in ecological studies, commercial kits have been increasingly utilized for DNA extraction and purification from soil and sediments [9,10,14,29-34].

The main goal of this study was to choose a method that would be inexpensive, able to process several samples quickly, and capable of obtaining high quality DNA for PCR studies. Four commercial DNA extraction kits and two physical-chemical methods (bead-beating and freeze-thaw) were compared for the extraction of DNA from a ferro-humic podzol soil amended or not with poultry manure and biochar. No earlier work suggested a mathematical decision aid tool to choose the optimum method for DNA extraction from soil under several conflicting criteria. The present work introduces a systematic/mathematical approach for comparing the performance of different DNA extraction methods under a set of simultaneous "multiple criteria": yield of extraction, purity of extracted DNA ( $A_{260/280}$  and  $A_{260/230}$  ratios), degradation degree of DNA, easiness of PCR amplification, duration of extraction, and cost per extraction. The proposed approach is formally called "multiple criteria decision-making" or MCDM [35].

In problems dealing with MCDM, which is a branch of Operations Research (or OR) models, the main goal is to consider a set of decision criteria and choose the best performing option from a list of available alternatives (*i.e.*, options to choose from), which generally show no obvious dominance one over another with respect to the criteria (this is formally referred to as Pareto optimality or Pareto efficiency condition). More precisely, it is assumed that all given alternatives (*e.g.*, here different DNA extraction techniques) are feasible and there is always a trade-off in choosing one over another. In other words, option A may be better than B under some criteria, but worse under some other criteria. In such complex decision-making scenarios, MCDM can aid the analyst (the decision maker) to make a final decision considering his/her experience, expectations, constraints, etc., into a systematic mathematical model [35,36]. It is worth adding that next to selection problems, there are also

MCDM models that are applicable to the sorting and classification problems [37]. The application of MCDM in decision-making processes in molecular biological systems, and more especially in comparing DNA extraction kits, is rather new. Most recently, a basic MCDM model, called "Weighted Sum Method" (WSM) was used for comparing a set of sample preservation and DNA extraction methods from swine feces [38]. Some other example applications of MCDM in a diverse range of practical problems include the use of decision analysis in integrated manufacturing [39], in the evaluation of technology investment decisions [40], in sustainable energy planning [41], and in prioritizing urban cultural heritage values [42].

Despite the fact that WSM is known to be the earliest, the simplest and probably the most widely used MCDM method, it has some shortcomings in particular decision-making cases. Namely, it allows a direct trade-off (compensation) among the criteria values in evaluating the performance of each alternative. As a result, it may choose an alternative that is excellent under some criteria but at the same time poor or close to unacceptably under some other criteria, which in turn can induce a risk in practice, especially under uncertain data or ambiguous conditions. For instance, a biologist may choose a DNA extraction kit using WSM that gives a very high yield but at the same time the purity of extracted DNA may be at a marginal level, which in turn can pose a primary risk/concern for that particular decision maker in terms of the quality of DNA. In turn, more advanced MCDM methods have been developed over years, among which is the TOPSIS (the Technique for Order Preference by Similarity to Ideal Solution) method developed by Hwang and Yoon [43]. It carries several advantages such as "simplicity, rationality, comprehensibility, good computational efficiency and ability to measure the relative performance for each alternative in a simple mathematical form" [44]. Most importantly, compared to other simple MCDM methods such as WSM, TOPSIS respects the fact that the decision maker sometimes likes to make as much profit as possible, but also to avoid as much risk as possible. The latter desire is satisfied with TOPSIS by introducing the concept of the ideal and negative-ideal (nadir) solutions. A selected alternative by this method should have the shortest distance from the ideal solution and the farthest distance from the negative-ideal solution in a geometrical sense [43]. The mathematical framework and application of this method in the context of the optimum DNA extraction selection method for agricultural soils are presented in a later section of the present article, which is the main motivation of the work. It should be added that the method is general enough to be applied to other decision-making processes in biological systems.

## 2. MATERIALS AND METHODS

### 2.1. Sample Collection and Preparation

Soil samples were collected in September 2011 at the Emile A. Lods Agronomy Research Centre of the Macdonald Campus Farm at McGill University, Quebec, Canada, in the range of 5 - 10 cm depth of A1 horizon (plowed soil). Collected samples were St-Amable sandy (Ferro-Humic Podzol) soil with the physical-chemical characteristics described in **Table 1**. Soil was air dried at room temperature to 10% - 20% water-holding capacity.

After sieving (stainless soil sieve with 2-mm mesh size), soil samples were stored at 4°C in the dark until analysis. Poultry manure was collected from the Poultry Complex of the Macdonald Campus Farm and subsequently air dried at room temperature, homogenized and stored at 4°C until analysis. Biochar used in this study

was obtained from wood lumbers and wastes by slow pyrolysis at 450°C (BlueLeaf Inc., Drummondville, QC, Canada). Mixtures were formulated with soil:manure [SM, 99:1(w/w)] and soil:manure:biochar [SMB, 98:1:1 (w/w)] and stored at 4°C in the dark until analysis.

### 2.2. DNA Extraction

Total bacterial DNA was extracted from soil, soil:manure [SM, 99:1(w/w)] and soil:manure:biochar [SMB, 98:1:1 (w/w)] mixtures using six different extraction methods (**Table 2**). The latter included two previously described home-made methods [38], as summarized below (a bead-beating technique and a freeze-thaw technique), as well as four commercial kits, as described by the manufacturers: PowerSoil<sup>®</sup> DNA Isolation Kit (MoBio Laboratories, Inc., Carlsbad, CA, USA); UltraClean<sup>™</sup> Soil

**Table 1.** Physical and chemical characteristics of soil.

Soil type	Sand (%)	Silt (%)	pH	Bulk density (Mg·m <sup>-3</sup> )	Organic matter (%)	Cation exchange capacity (cmol·kg <sup>-1</sup> )	Hydraulic conductivity (cm·d <sup>-1</sup> )
Sandy	92.2	4.3	5.5	1.350	2.97	4.9	1.67 ± 0.45

**Table 2.** Comparative analysis of DNA extraction techniques used in this study.

Parameters	Commercial kits <sup>†</sup>								Physical-Chemical techniques	
	PowerSoil		UltraClean		FastSPIN		E.Z.N.A.		Bead-Beating Freeze-Thaw	
	conventional	alternative	conventional	alternative	conventional	alternative	conventional	alternative		
Sample weight (mg)	250		250		500		500		100	100
Beads	Unknown		Unknown		Unknown		Glass		Glass	None
Cell lysis solution	Buffer contains sodium dodecyl sulphate (SDS)				SDS Buffer, MT Buffer		Buffer of SLX Mlus		SDS and proteinase K	
Cell lysis technique	Vortex at maximum speed for 10 min.		Vortex 3 - 4 sec. then heating at 70° for 5 min.		Vortex 3 - 4 sec. then heating at 70° for 5 min.		FastPrep <sup>®</sup> Instrument for 40 sec. at speed setting of 6.0		Thermal shock	
Protein removal	Patented Inhibitor Removal Technology		Solution S2		Protein Precipitation Solution		Buffer XP1		PCI <sup>‡</sup>	
Humic acid removal	Patented Inhibitor Removal Technology		Inhibitor Removal Solution		None		HTR Reagent		None	
DNA precipitation	High concentration salt solution				Binding Matrix		Isopropanol		Polyethyleneglycol and isopropanol	
DNA purification	Spin filter with silica membrane				SPIN Filter		HiBind DNA column		CI <sup>§</sup>	
Elution buffer	10 mM Tris-HCl, pH 8.0		10 mM Tris-HCl, pH 8.0		DNase/Pyrogen-Free Water		Unknown		T <sub>1</sub> E <sub>0.1</sub> buffer: 1 mM Tris-HCl, 0.1 mM EDTA, pH 8.0	
Final volume (µl)	100		50		None		Incubation with Binding Matrix for 5 min at 55°C		50	

<sup>†</sup>PowerSoil: PowerSoil<sup>®</sup> DNA Isolation Kit; UltraClean: UltraClean<sup>™</sup> Soil DNA Isolation Kit; FastSPIN: FastDNA<sup>®</sup> SPIN Kit for Soil; E.Z.N.A.: E.Z.N.A.<sup>®</sup> Soil DNA Isolation Kit. PCI<sup>‡</sup>: phenol:chloroform:isoamyl alcohol, 25:24:1 (v/v); CI<sup>§</sup>: chloroform:isoamyl alcohol, 24:1 (v/v).



DNA Isolation Kit (MoBio Laboratories, Inc., Carlsbad, CA, USA); FastDNA<sup>®</sup> SPIN Kit for Soil (MP Biomedicals, LLC, Solon, OH, USA); E.Z.N.A.<sup>®</sup> Soil DNA Isolation Kit (Omega Bio-Tek, Inc., Carlsbad, CA, USA). Next to the conventional procedure for each of the four DNA extraction commercial kits, to reduce DNA shearing and/or increase DNA yields, alternative lysis methods were assessed as described by the manufacturers (**Table 2**), hence resulting in a total of 10 methods including the home-made methods.

### Bead-Beating and Freeze-Thaw Methods

Soil, soil:manure [SM, 99:1(w/w)] and soil:manure:biochar [SMB, 98:1:1(w/w)] mixtures (0.1 g) were resuspended in 1 ml of extraction buffer (500 mM Tris-HCl pH 8.0, 100 mM sodium EDTA pH 8.0, 1.5 M NaCl) and homogenized by vortexing. For disrupting cells by bead-beating, 0.1 g of 0.1 mm-diameter glass beads (BioSpec Products, Bartlesville, OK, USA) were added and cells were disrupted by shaking the tubes for 40 sec (speed = 6.0) in a Fast-Prep (Bio101 Fast-Prep model FP120, Thermo Savant, Qiagen, Inc., Carlsbad, CA, USA), leaving on ice for 5 min (to counteract heating of the tubes in the Fast-Prep), and shaking a second time for 40 sec (speed = 6.0). For disrupting cells by freeze-thaw, three cycles of freezing in liquid nitrogen (−196°C) for 5 min and thawing at 65°C in a water bath for 10 min were used.

Subsequent procedures for DNA purification after both bead-beating and freeze-thaw were the same. After physical disruption of the cells either by bead-beating or freeze-thaw, 20 µl of proteinase K were added and tubes were incubated at 37°C with shaking at 180 rpm for 30 min to digest contaminating proteins. One hundred microliters of 20% (w/v) SDS (sodium dodecyl sulfate) were added, tubes were mixed by inverting several times, incubated at 65°C for 1 h and centrifuged at 12,000 × g for 5 min. The supernatants were transferred to new tubes containing 0.5 volume of polyethylene glycol (PEG) solution [30% (w/v) PEG, 1.5 M NaCl] and incubated at room temperature for 2 h. Tubes were centrifuged at 16,100 × g for 20 min. The pellets were dissolved in 90 µl of T<sub>1</sub>E<sub>0.1</sub> buffer (1 mM Tris-HCl, 0.1 mM EDTA, pH 8.0). Thirty microliters of ammonium acetate 10 M were added, tubes were mixed by inverting several times and left on ice for 5 min to counteract heating. Tubes were centrifuged at 16,100 × g for 30 min at 4°C to precipitate proteins and polysaccharides. The DNA was purified from the aqueous phase by phenolchloroform-isoamyl alcohol extraction [25:24:1 (v/v)] followed by chloroform-isoamyl alcohol extraction [24:1 (v/v)]. The DNA was precipitated by adding 0.6 volume of isopropanol and incubating at −20°C for 1 h. DNA was collected by centrifuging at 16,100 × g for 10 min at 4°C.

The pellets were washed twice with 70% ethanol. After air drying for about 30 min, the pellets were resuspended in 80 µl of T<sub>1</sub>E<sub>0.1</sub> buffer (1 mM Tris-HCl, 0.1 mM EDTA, pH 8.0). The resulting extracts were treated with 10 µg of RNase (Invitrogen, Burlington, ON, Canada) for 10 min at 37°C, and stored at −20°C.

### 2.3. Polymerase Chain Reaction Amplification

PCR amplifications were carried out in a Veriti<sup>™</sup> Thermal Cycler (Applied Biosystems, Foster City, CA, USA) using 10 ng of DNA extracted from each sample by each DNA extraction technique as template. The V3 region of the bacteria 16S rDNA was targeted using the Bacteria universal primers 341F (forward primer: CCTACGGGAGGCAGCAG) and 534R (reverse primer: ATTACCGCGGCTGCTGG), which yield amplicons of about 193 bp [45]. The PCR reaction mixture contained 0.75 µM of each primer, 200 µM of each dNTP (Amersham Biosciences Corp., Piscataway, NJ, USA), 1.25 U *Taq* DNA polymerase (Invitrogen, Burlington, ON, Canada), and the PCR buffer supplied with the enzyme (10 mM Tris-HCl pH 9.0, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>) [46]. For each DNA extract, the following series of PCR tubes were analyzed for the presence of the V3 region of the bacteria 16S rDNA: 1) triplicate PCR tubes with 10 ng of extracted DNA; 2) a positive control tube with 10 ng of DNA extracted from a pure culture of *Escherichia coli* ATCC 25922; 3) an inhibition control tube with 5 ng of DNA extracted from a pure culture of *E. coli* ATCC 25922 and 5 ng of DNA extracted from each sample, in order to assess the presence of PCR inhibitors in the extracts; 4) a negative control tube consisting of the reaction mixture without DNA, in order to assess the presence of external or cross-contamination of the PCR reaction mixtures by DNA.

The PCR conditions were 5 min at 99°C (initial denaturation), then 2 cycles of 5 min at 94°C (denaturation), 5 min at 55°C (annealing) and 2 min at 72°C (extension), then 28 cycles of 1 min at 94°C (denaturation), 1 min at 55°C (annealing), 2 min at 72°C (extension), and finally an extension period of 10 min at 72°C. The size (about 193 bp), specificity (unique band), and abundance of PCR products were determined by comparison with DNA standards (GeneRuler 100 bp DNA Ladder, MBI Fermentas, Burlington, ON, Canada) after agarose gel electrophoresis [46].

### 2.4. Identifying a Set of Criteria for Comparing DNA Extraction Methods

All DNA extractions were performed in 5 replicates. For each of the following criteria, errors are indicated in **Table 3** as the standard deviation of 3 replicate measurements on 5 replicate extractions ( $n = 15$ ).



**Table 3.** Performance of each DNA extraction method under seven decision criteria for soil, SM [soil:manure, 99:1(w/w)] and SMB [soil:manure:biochar, 98:1:1(w/w)].

Methods <sup>†</sup>	Sample	Performance/Decision Criteria						
		C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>	C <sub>5</sub>	C <sub>6</sub>	C <sub>7</sub>
		Yield (µg DNA/g sample)	A <sub>260/280</sub> ratio	A <sub>260/230</sub> ratio	Degree of DNA degradation <sup>§</sup>	Easiness of amplification <sup>#</sup>	Duration of extraction (hrs.)	Cost per extraction (CAD) <sup>††</sup>
UltraClean Conventional	Soil	8.21 ± 1.32 <sup>‡</sup>	1.75 ± 0.02	1.30 ± 0.05	2	2	0.83	4.64
	SM	14.21 ± 1.24	1.71 ± 0.06	1.43 ± 0.20	2	2	0.83	4.64
	SMB	13.03 ± 1.23	1.77 ± 0.04	1.21 ± 0.30	2	2	0.83	4.64
UltraClean Alternative	Soil	5.59 ± 0.58	1.63 ± 0.23	1.22 ± 0.06	1	2	0.92	4.64
	SM	10.77 ± 1.39	1.50 ± 0.04	0.81 ± 0.04	1	2	0.92	4.64
	SMB	9.40 ± 1.36	1.53 ± 0.06	0.88 ± 0.03	1	2	0.92	4.64
PowerSoil Conventional	Soil	6.76 ± 0.84	1.67 ± 0.08	1.28 ± 0.18	1	2	1.00	5.54
	SM	10.52 ± 1.69	1.96 ± 0.05	1.67 ± 0.13	1	2	1.00	5.54
	SMB	8.75 ± 1.43	1.90 ± 0.07	1.57 ± 0.15	1	2	1.00	5.54
PowerSoil Alternative	Soil	3.78 ± 0.32	1.74 ± 0.34	2.07 ± 0.10	1	2	1.08	5.54
	SM	10.48 ± 3.58	1.55 ± 0.06	0.86 ± 0.11	1	2	1.08	5.54
	SMB	11.97 ± 2.17	1.55 ± 0.05	0.96 ± 0.11	1	2	1.08	5.54
FastSpin Conventional	Soil	17.70 ± 1.72	1.73 ± 0.02	0.25 ± 0.02	3	1	1.16	6.58
	SM	20.00 ± 1.64	1.77 ± 0.03	0.39 ± 0.04	3	1	1.16	6.58
	SMB	23.87 ± 2.60	1.84 ± 0.02	0.60 ± 0.07	3	1	1.16	6.58
FastSpin Alternative	Soil	23.29 ± 1.54	1.70 ± 0.02	0.34 ± 0.12	3	1	1.25	6.58
	SM	20.36 ± 0.88	1.77 ± 0.04	0.50 ± 0.09	3	1	1.25	6.58
	SMB	21.45 ± 1.95	1.79 ± 0.01	0.47 ± 0.01	3	1	1.25	6.58
E.Z.N.A. Conventional	Soil	4.33 ± 0.18	1.67 ± 0.08	0.43 ± 0.42	3	1	3.17	4.38
	SM	7.26 ± 0.1.30	1.85 ± 0.06	1.57 ± 0.28	3	1	3.17	4.38
	SMB	7.69 ± 0.1.32	1.87 ± 0.06	1.70 ± 0.16	3	1	3.17	4.38
E.Z.N.A. Alternative	Soil	25.98 ± 4.86	1.56 ± 0.02	0.30 ± 0.20	3	1	3.25	4.38
	SM	25.90 ± 2.92	1.57 ± 0.04	0.55 ± 0.20	3	1	3.25	4.38
	SMB	15.73 ± 3.53	1.65 ± 0.01	0.47 ± 0.33	3	1	3.25	4.38
Bead-Beating	Soil	5.24 ± 1.03	0.77 ± 0.05	0.12 ± 0.01	ND <sup>¶</sup>	3	7.5	1
	SM	6.53 ± 0.36	0.87 ± 0.06	0.11 ± 0.01	ND	3	7.5	1
	SMB	4.87 ± 0.44	0.78 ± 0.09	0.09 ± 0.01	ND	3	7.5	1
Freeze-Thaw	Soil	5.49 ± 1.10	1.13 ± 0.18	0.19 ± 0.07	ND	3	8.5	1
	SM	6.21 ± 0.64	0.85 ± 0.07	0.11 ± 0.01	ND	3	8.5	1
	SMB	8.90 ± 2.73	0.86 ± 0.06	0.13 ± 0.02	ND	3	8.5	1

<sup>†</sup>UltraClean: UltraClean™ Soil DNA Isolation Kit; PowerSoil: PowerSoil® DNA Isolation Kit; FastSPIN: FastDNA® SPIN Kit for Soil; E.Z.N.A.: E.Z.N.A.® Soil DNA Isolation Kit. PCI; <sup>‡</sup>: Each test was repeated five times and the ± values refer to their standard deviations. <sup>§</sup>Degree of DNA degradation (Lemarchand *et al.*, 2005): 1 = low (mean fragment size between 23 and 2 kb); 2 = medium (mean fragment size between 23 and 0.5 kb); 3 = high (mean fragment size between 23 and < 0.5 kb). <sup>#</sup>1 = low (mean fragment size between 23 and 2 kb); 2 = medium (mean fragment size between 23 and 0.5 kb); 3 = high (mean fragment size between 23 and <0.5 kb); <sup>††</sup>Canadian dollars; <sup>¶</sup>Not determined since extracted DNA was not visible on the agarose gel stained with ethidium bromide.

DNA concentration ( $\text{ng DNA}\cdot\mu\text{l}^{-1}$ ),  $A_{260/280}$  ratio (absorbance at 260 nm/absorbance at 280 nm) and  $A_{260/230}$  ratio (absorbance at 260 nm/absorbance at 230 nm) of each extract were determined using a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, Marietta, OH, USA). The yield for each DNA extraction method was calculated as follows: Yield of extraction ( $\mu\text{g}$  of DNA/g of sample) = concentration of DNA in the extract ( $\text{ng}/\mu\text{l}$ )  $\times$  (1  $\mu\text{g}/1000$  ng)  $\times$  final volume of extract ( $\mu\text{l}$ )/dry weight of sample (g).

The  $A_{260/280}$  ratio and the  $A_{260/230}$  ratio were used to evaluate the purity of DNA extracts. An  $A_{260/280}$  ratio higher than 1.8 indicates the absence of proteins in DNA extracts. When the  $A_{260/280}$  ratio is lower than 1.8, proteins or other contaminants (co-extracted with DNA) that absorb strongly at or near 280 nm may be present. An  $A_{260/280}$  ratio over 2.0 indicates RNA contamination of the sample.

An  $A_{260/230}$  ratio between 2.0 and 2.2 is indicative of the high purity of extracted DNA. When the  $A_{260/230}$  ratio is lower than 2, humic acids, carbohydrates, phenol, guanidine HCl or other contaminants that absorb at or near 230 nm, may be presented [47].

The DNA in each extract was checked for integrity (degradation degree) by agarose gel electrophoresis by comparing with Lambda DNA HindIII Digest standards (New England BioLabs, Ipswich, MA, USA) using AlphaEaseFC software version 3.1.2 (AlphaInnotech Corporation, San Leandro, CA, USA). The degradation degree of the DNA in each extract was evaluated using the scale proposed by Lemarchand *et al.* (2005): 1 = low (mean fragment size between 23 and 2 kb); 2 = medium (mean fragment size between 23 and 0.5 kb); 3 = high (mean fragment size between 23 and <0.5 kb).

In addition to the above, in this study we used a new criterion "Easiness of amplification" for comparing the ten DNA extraction methods (Table 2). The number of PCR bands as well as the presence or absence of *Taq* DNA polymerase inhibitors dictated the "Easiness of amplification", which was expressed by a quantitative scheme as follows: [1 = easy to perform (one band detected, absence of inhibitors), 2 = moderately difficult to perform (1 band detected and presence of inhibitors, or 2 bands detected and absence of inhibitors), 3 = very difficult to perform (2 bands detected and presence of inhibitors, or no band detected)].

Two other criteria, namely duration of extraction and cost per extraction, were included for a total of seven decision criteria to compare the ten different DNA extraction methods and choose the best one via a systematic mathematical method as follows.

## 2.5. Multiple Criteria Decision-Making (MCDM): The Entropy and TOPSIS Techniques

A decision-making process often involves making pref-

erence decisions over multitude alternatives (given options) that are characterized by multiple, usually conflicting criteria (Ahn, 2011). A typical decision matrix,  $X$ , used in MCDM is shown in Figure 1, where  $C_j$  represent the decision criteria ( $j = 1, \dots, n$ );  $A_i$  represent the alternatives ( $i = 1, \dots, m$ ), and  $x_{ij}$  represent the value of the  $i$ -th alternative under the  $j$ -th criterion.  $W_j$  ( $j = 1, \dots, n$ ) are the criteria weights, indicating the relative importance among them.

Among different criteria weight assignment techniques used in the MCDM field, the "Entropy" method is among the very few techniques that are independent of the decision maker's subjective priorities/judgments. In the entropy method, the criteria weights are calculated based on the actual measured data in the decision matrix; *i.e.*, by means of an objective/statistical process.

Following Section 2.4 in the present study, the ten different DNA extraction methods (alternatives) are to be compared under seven performance criteria ( $C_j$ ,  $j = 1, 2, \dots, 7$ ) including: ( $C_1$ ) yield of extraction (the higher the better); ( $C_2$ )  $A_{260/280}$  ratio (the higher the better); ( $C_3$ )  $A_{260/230}$  ratio (the higher the better); ( $C_4$ ) degree of DNA degradation (the lower the better); ( $C_5$ ) easiness of amplification ranking index (the lower the better); ( $C_6$ ) duration of extraction (the lower the better); and ( $C_7$ ) cost per extraction (the lower the better). Decision weights using the entropy technique were calculated for all criteria via the following steps [48].

### Step 1. Transferring the decision matrix to the normalized mode.

In order to adjust the entropy measure for the  $j$ -th criterion, the related values in the decision matrix are first normalized as  $P_{ij}$ ;

$$P_{ij} = \frac{x_{ij}}{\sum_i x_{ij}} \quad (1)$$

### Step 2. Calculating the entropy of dataset for each criterion.

In this step, the entropy of the  $j$ -th criterion,  $E_j$ , is calculated as follows:

$$E_j = -a \sum_{i=1}^m p_{ij} \ln p_{ij} \quad (2)$$

$(i = 1, 2, \dots, m), (j = 1, 2, \dots, n)$

criteria	$C_1$	$C_2$	$\dots$	$C_n$
(weights	$w_1$	$w_2$	$\dots$	$w_n$ )
alternatives	-----			
$A_1$	$x_{11}$	$x_{12}$	$\dots$	$x_{1n}$
$A_2$	$x_{21}$	$x_{22}$	$\dots$	$x_{2n}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$A_m$	$x_{m1}$	$x_{m2}$	$\dots$	$x_{mn}$

Figure 1. A typical decision matrix in MCDM.

where,  $\alpha = 1/\ln(m)$ ; “ $m$ ” is the total number of alternatives (in this study, the DNA extraction methods over different samples);

Next, the operation of subtraction is used to measure the degree of diversity relative to the corresponding anchor value (unity),  $D_j$ , using the following formula:

$$D_j = 1 - E_j \quad (3)$$

### Step 3. Defining criteria weights.

The entropy weight of each criterion is calculated using:

$$W_j = D_j / \sum_{i=1}^m D_j \quad (4)$$

These weights are then incorporated into the so-called TOPSIS MCDM technique to calculate an overall score for each DNA extraction method. The TOPSIS technique was chosen because of its high speed, accuracy, and compatibility [49]. The algorithm of this technique is summarized as follows:

1) Transfer the decision matrix to the normalized mode;

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m x_{ij}^2}} (i = 1, 2, \dots, m), (j = 1, 2, \dots, n) \quad (5)$$

2) Weigh the normalized decision matrix;

$$v_{ij} = W_j \times r_{ij} (i = 1, 2, \dots, m), (j = 1, 2, \dots, n) \quad (6)$$

3) Define the “ideal positive”  $V_j^+$  and “ideal negative (nadir)”  $V_j^-$  solutions;

$$\begin{cases} \{V_1^+, V_2^+, \dots, V_n^+\} = \{(\max_i V_{ij} | j \in J), \\ (\min_i V_{ij} | j \in J') | i = 1, 2, \dots, m\} \\ \{V_1^-, V_2^-, \dots, V_n^-\} = \{(\max_i V_{ij} | j \in J), \\ (\max_i V_{ij} | j \in J') | i = 1, 2, \dots, m\} \end{cases} \quad (7)$$

4) Measure the distances,  $d_i^+$  and  $d_i^-$ ,  $i = 1, 2, \dots, n$ , from the ideal and negative ideal solutions;

$$\begin{cases} d_i^+ = \left[ \sum_{i=1}^n (V_{ij} - V_j^+)^2 \right]^{1/2}, i = 1, 2, \dots, m \\ d_i^- = \left[ \sum_{i=1}^n (V_{ij} - V_j^-)^2 \right]^{1/2}, i = 1, 2, \dots, m \end{cases} \quad (8)$$

5) Determine the relative closeness of alternatives to the ideal solution;

$$C_i^+ = \frac{d_i^-}{d_i^+ + d_i^-}, i = 1, 2, \dots, m \quad (9)$$

where  $0 \leq C_i^+ \leq 1$ . Alternatives with higher magnitudes of closeness  $C_i^+$  are more preferred.

## 3. RESULTS

### 3.1. Revealing Conflicts among the Selection Criteria: A One-Criterion-at-a-Time Analysis

The performance of the ten DNA extraction methods under the seven decision criteria for soil, SM [soil:manure, 99:1(w/w)] and SMB [soil:manure:biochar, 98:1:1 (w/w)] are displayed in **Table 3** (which in fact can be considered as the given decision matrix in the MCDM terminology). Altogether, 30 different DNA extracts were obtained (10 extraction methods for each of the three sample types). Subsequently, using the values presented in **Table 3**, the 10 extraction methods were evaluated by seven decision criteria for each of the three sample types and were ranked in **Table 4**. The ranking results in **Table 4** show that the E.Z.N.A alternative extraction method provided the highest yield of extraction for soil and SM samples, whereas for SMB samples, the FastSPIN conventional method had the best performance in terms of yield of extraction compared to the other commercial kits and home-made physical-chemical techniques.

Under the purity (ratio  $A_{260/280}$  and  $A_{260/230}$ ) criteria, the PowerSoil conventional and E.Z.N.A. conventional methods ranked 1<sup>st</sup> and 2<sup>nd</sup>, respectively, for SM and SMB samples (**Table 4**). In contrast, under the  $A_{260/280}$  criterion for soil samples, the UltraClean conventional method gained the first rank, followed by the PowerSoil alternative method. However, under the  $A_{260/230}$  criterion for soil samples, the latter method gained the first rank and the UltraClean conventional method received the second rank.

The degree of degradation of extracted DNA also varied depending on the extraction technique applied (**Table 3**). It was the highest when the conventional and alternative lysis methods of the FastSPIN and E.Z.N.A. kits were used (**Table 4**), and was the lowest using the PowerSoil conventional and alternative methods as well as the UltraClean alternative method. Among all commercial kits, it is also worth noticing from **Table 3** that the UltraClean kit was the only one for which the alternative method reduced the degree of DNA degradation.

Under the “Easiness of amplification” criterion, consistent specific PCR amplification (unique band) of ~193 bp amplicons corresponding to the V3 region of the Bacteria 16S rDNA was successfully obtained for DNA extracted from soil, SM and SMB mixtures using both conventional and alternative methods of FastSPIN and E.Z.N.A. (results not shown): hence they resulted in the top rank under this criterion (**Table 4**). In contrast, two bands were detected after amplification of DNA extracted from soil, SM and SMB mixtures using both conventional and alternative methods of UltraClean and PowerSoil commercial kits (results not shown). In addi-

**Table 4.** Ranks of DNA extraction methods for seven decision criteria; each group of three numbers separated with commas indicate: soil, SM [soil:manure, 99:1(w/w)] and SMB [soil:manure:biochar, 98:1:1(w/w)].

Methods <sup>†</sup>	Ranks in decision criteria <sup>‡</sup>						
	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>	C <sub>5</sub>	C <sub>6</sub>	C <sub>7</sub>
	Yield of extraction	A <sub>260/280</sub> ratio	A <sub>260/230</sub> ratio	Degree of DNA degradation	Easiness of amplification	Duration of extraction	Cost per extraction
UltraClean Conventional	4, 4, 4	1, 5, 5	2, 3, 3	2, 2, 2	2, 2, 2	1, 1, 1	3, 3, 3
UltraClean Alternative	6, 5, 6	7, 8, 8	4, 5, 5	1, 1, 1	2, 2, 2	2, 2, 2	3, 3, 3
PowerSoil Conventional	5, 6, 8	5, 1, 1	3, 1, 1	1, 1, 1	2, 2, 2	3, 3, 3	4, 4, 4
PowerSoil Alternative	10, 7, 5	2, 7, 7	1, 4, 4	1, 1, 1	2, 2, 2	4, 4, 4	4, 4, 4
FastSpin Conventional	3, 3, 1	3, 3, 3	8, 8, 6	3, 3, 3	1, 1, 1	5, 5, 5	5, 5, 5
FastSpin Alternative	2, 2, 2	4, 4, 4	7, 7, 8	3, 3, 3	1, 1, 1	6, 6, 6	5, 5, 5
E.Z.N.A. Conventional	9, 8, 9	6, 2, 2	6, 2, 2	3, 3, 3	1, 1, 1	7, 7, 7	2, 2, 2
E.Z.N.A. Alternative	1, 1, 3	8, 6, 6	5, 6, 7	3, 3, 3	1, 1, 1	8, 8, 8	2, 2, 2
Bead-Beating	8, 9, 10	10, 9, 10	10, 9, 10	ND*	3, 3, 3	9, 9, 9	1, 1, 1
Freeze-Thaw	7, 10, 7	9, 10, 9	9, 10, 9	ND*	3, 3, 3	10, 10, 10	1, 1, 1

<sup>†</sup>UltraClean: UltraClean™ Soil DNA Isolation Kit; PowerSoil: PowerSoil® DNA Isolation Kit; FastSPIN: FastDNA® SPIN Kit for Soil; E.Z.N.A.: E.Z.N.A.® Soil DNA Isolation Kit. <sup>‡</sup>Rank of each DNA extraction method for soil, soil:manure, and soil:manure:biochar, respectively, under each decision criterion. \*: Not determined.

tion, no band was detected after amplification of DNA extracted by the Bead-Beating and Freeze-Thaw techniques. These results imply a requirement for optimization of the PCR reaction mixture and/or program, as well as the necessity for assessing the presence of *Taq* DNA polymerase inhibitor(s).

For DNA extracted by both conventional and alternative methods of the four commercial kits as well as home-made methods (Bead-Beating and Freeze-Thaw techniques), the specific amplification of ~193 bp amplicons was obtained for each extract in the inhibition controls (mixture of DNA extracted from *E. coli* ATCC 25922 and DNA extracted from each sample), indicating the absence of *Taq* DNA polymerase inhibitors in the extracts from these DNA extraction techniques. The absence of PCR band is in line with the low yields and purities of DNA extracted with these two techniques (Table 3), making them hold the lowest (3<sup>rd</sup>) rank under the easiness of amplification criterion.

The duration of extraction and the cost per extraction are also very important in selecting an optimum DNA extraction technique in view of processing a large number of environmental samples in ecological studies. Under the duration of extraction criterion, UltraClean conventional method obtained the first rank, followed by the alternative type of UltraClean method (Table 4). The home-made Bead-Beating and Freeze-Thaw techniques performed the poorest under all the criteria except for the cost where they were ranked as the preferred options (least costly). Since the DNA extracted with the Bead-Beating and Freeze-Thaw methods was not visible on the agarose gel stained with ethidium bromide because of very low concentrations (6.08 and 8.16 ng of DNA/μl of

extract, respectively, Table 3), the degree of DNA degradation could not be determined.

The above comparison of the DNA extraction methods under each criterion individually (Table 4) clearly demonstrates the presence of conflicts among criteria in the given decision-making problem. An example of such conflict is for the FastSpin conventional extraction method which resulted in a good performance in terms of yield of DNA extraction from soil, SM, and SMB (ranks of 3, 3, and 1 respectively), but performed poorly under the degree of DNA degradation criterion. In contrast, the DNA extracted by the E.Z.N.A. conventional method offered a good performance under Easiness of amplification (rank 1 for all three sample types), but its yield of extraction was one of the lowest (ranks of 9, 8, and 9 for soil, SM, and SMB, respectively). Because of such conflicts, the MCDM Entropy-TOPSIS approach was deemed necessary and implemented in order to choose overall the best DNA extraction method under simultaneous decision-making criteria for each specific soil mixture.

### 3.2. Multiple Criteria Decision-Making (MCDM): The Entropy Method

Following Section 2.5, as the first step to the MCDM solution, criteria importance weights needed to be calculated for all seven criteria for soil, SM and SMB using the entropy method according to Formulas 1-4. For example, to calculate the weight of criterion C<sub>1</sub> for DNA extraction from soil, the normalization was first performed to calculate  $p_{ij}$  values using formula 1 (Table 5). Then,  $p_{ij} \times \ln(p_{ij})$  values were calculated (see the exam-

**Table 5.** Normalized decision matrix for DNA extraction methods for soil.

Methods <sup>†</sup>	Normalized Decision Matrix Data (p <sub>ij</sub> )						
	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>	C <sub>5</sub>	C <sub>6</sub>	C <sub>7</sub>
	Yield of extraction	A <sub>260/280</sub> ratio	A <sub>260/230</sub> ratio	Degree of DNA degradation	Easiness of amplification	Duration of extraction	Cost per extraction
UltraClean Conventional	0.09 <sup>‡</sup>	0.13	0.18	0.12	0.10	0.07	0.11
UltraClean Alternative	0.06	0.12	0.16	0.06	0.10	0.07	0.11
PowerSoil Conventional	0.07	0.12	0.17	0.06	0.10	0.08	0.13
PowerSoil Alternative	0.04	0.13	0.28	0.06	0.10	0.09	0.13
FastSpin Conventional	0.19	0.13	0.03	0.18	0.15	0.09	0.16
FastSpin Alternative	0.24	0.13	0.05	0.18	0.15	0.10	0.16
E.Z.N.A. Conventional	0.05	0.12	0.06	0.18	0.15	0.25	0.10
E.Z.N.A. Alternative	0.27	0.12	0.07	0.18	0.15	0.26	0.10

<sup>†</sup>UltraClean: UltraClean™ Soil DNA Isolation Kit; PowerSoil: PowerSoil® DNA Isolation Kit; FastSPIN: FastDNA® SPIN Kit for Soil; E.Z.N.A.: E.Z.N.A.® Soil DNA Isolation Kit. <sup>‡</sup>Example of calculation: Yield of extraction criterion for soil (using **Table 3**):  $0.09 = 8.21 / (8.21 + 5.59 + 6.76 + 3.78 + 17.70 + 23.29 + 4.33 + 25.98)$ .

ple of calculation in **Table 6**) followed by the calculation of  $E_j$  and  $D_j$  using Formulas 2 and 3, respectively (**Table 6**). Finally the weight of each criterion was calculated using Formula 4. All criteria weights for soil, SM and SMB are summarized in **Table 7**. It should be added that in some cases the decision maker is experienced enough to have his/her own (subjective) weights, which can be combined by the (objective) Entropy weights extracted from Equation (4). The mathematical framework for the latter combined weighting scheme can be found in [43]. Here it is assumed that the decision maker is inexperienced and/or conservative where he/she has an equal preference towards the performance criteria and hence prefers to purely rely on the Entropy weights.

### 3.3. Selecting an Optimal DNA Extraction Method for Each Sample Type: The TOPSIS Method

The obtained weights of criteria were incorporated into the TOPSIS technique (Formulas 5-9) to calculate an overall score for each DNA extraction method for soil, SM, and SMB (**Tables 8-12**). The final rankings of the extraction methods for different sample types are shown in **Tables 10-12**. Based on these results, the extraction methods were ranked as follows:

For soil (**Table 10**, descending order): PowerSoil Alternative > FastSpin Alternative > E.Z.N.A. Alternative > UltraClean Conventional > PowerSoil Conventional > UltraClean Alternative > FastSpin Conventional > E.Z.N.A. Conventional > Bead-Beating = Freeze-Thaw;

For SM (**Table 11**, descending order): PowerSoil Conventional > UltraClean Conventional > UltraClean Alternative > PowerSoil Alternative > FastSpin Alternative

> FastSpin Conventional > E.Z.N.A. Conventional > E.Z.N.A. Alternative > Bead-Beating = Freeze-Thaw;

For SMB (**Table 12**, descending order): PowerSoil Conventional > UltraClean Conventional > PowerSoil Alternative > UltraClean Alternative > FastSpin Conventional > FastSpin Alternative > E.Z.N.A. Conventional > E.Z.N.A. Alternative > Bead-Beating = Freeze-Thaw.

## 4. DISCUSSION

The selection of an appropriate method for extracting DNA from complex ecosystems such as soil has a critical impact on the composition and richness of detected microbial communities using culture-independent molecular microbiological methods, such as competitive PCR, real-time PCR, denaturing gradient gel electrophoresis (DGGE) and Next-Generation DNA Sequencing (NGS) technologies [9,10,11,29]. In addition, it has been demonstrated that the quality of extracted DNA can interfere with microarray hybridizations, yielding high background noise and false positives [9,10,24,50]. These arguments show the necessity for carefully selecting a suitable DNA extraction method for each given sample type and proposed downstream DNA-based analyses.

In the present study, we compared both the quantity and quality of DNA extracted from soil, SM and SMB mixtures using 10 different DNA extraction methods to be used subsequently for PCR analysis. The results showed that both the quantity (yield of extraction) and the quality (purity, degree of degradation, easiness of amplification) of the extracted DNA depended on the extraction method and the type of environmental sample (**Table 3**), which was in agreement with earlier observations reported in the literature [9,10,14,32,33]. Moreover,



**Table 6.** Calculating the entropy of data (column-wise) for each decision criterion for soil.

Methods <sup>†</sup>	$p_{ij} \times \ln p_{ij}$						
	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$
	Yield of extraction	A <sub>260/280</sub> ratio	A <sub>260/230</sub> ratio	Degree of DNA degradation	Easiness of amplification	Duration of extraction	Cost per extraction
UltraClean Conventional	-0.21 <sup>‡</sup>	-0.27	-0.31	-0.25	-0.23	-0.18	-0.24
UltraClean Alternative	-0.17	-0.26	-0.30	-0.17	-0.23	-0.19	-0.24
PowerSoil Conventional	-0.19	-0.26	-0.30	-0.17	-0.23	-0.20	-0.27
PowerSoil Alternative	-0.13	-0.26	-0.36	-0.17	-0.23	-0.21	-0.27
Fast Spin Conventional	-0.31	-0.26	-0.11	-0.31	-0.28	-0.22	-0.29
Fast Spin Alternative	-0.34	-0.26	-0.14	-0.31	-0.28	-0.23	-0.29
E.Z.N.A. Conventional	-0.14	-0.26	-0.16	-0.31	-0.28	-0.35	-0.23
E.Z.N.A. Alternative	-0.35	-0.25	-0.19	-0.31	-0.28	-0.35	-0.23
Sum	-1.84	-2.08	-1.87	-1.98	-2.06	-1.92	-2.07
$E_j^{\S}$	0.89 <sup>¶</sup>	1.00	0.90	0.95	0.99	0.92	0.99
$D_j^{\#}$	0.11 <sup>††</sup>	0.00	0.10	0.05	0.01	0.08	0.01

<sup>†</sup>UltraClean: UltraClean™ Soil DNA Isolation Kit; PowerSoil: PowerSoil® DNA Isolation Kit; FastSPIN: FastDNA® SPIN Kit for Soil; E.Z.N.A.: E.Z.N.A.® Soil DNA Isolation Kit. <sup>‡</sup>Example of calculation:  $p_{11} \times \ln p_{11}$  (using **Table 5**, Formula 1) =  $0.09 \times \ln(0.09)$ . <sup>§</sup> $E_j$  = entropy of the set of normalized data. <sup>¶</sup>Example of calculation of  $E_1$  where  $\alpha = 0.48$  (Formula 2):  $0.65 = -0.48 \times [(-0.21) + (-0.17) + (-0.19)] + (-0.13) + (-0.31) + (-0.34)$ . <sup>#</sup> $D_j$  = degree of diversity. <sup>††</sup>Example of calculation of  $D_1$ :  $0.35 = 1 - 0.65$  (Formula 3).

**Table 7.** Criteria weights ( $W_{j,j} = 1, 2, \dots, 7$ ) for soil, SM [soil:manure, 99:1(w/w)] and SMB [soil:manure:biochar, 98:1:1(w/w)].

	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$
Soil	0.32 <sup>†</sup>	0.00	0.28	0.14	0.03	0.21	0.02
SM	0.16	0.01	0.25	0.21	0.04	0.31	0.03
SMB	0.16	0.01	0.22	0.22	0.04	0.33	0.03

<sup>†</sup>Example of calculation of  $W_1$  (weight of yield of extraction for soil, formula 4), using  $D_1$  in **Table 6**:  $0.32 = 0.11/(0.11 + 0.00 + 0.10 + 0.05 + 0.01 + 0.08 + 0.01)$ .

**Table 8.** Summary of normalized decision matrix data for TOPSIS method using seven decision criteria for each DNA extraction method for soil.

Methods <sup>†</sup>	Normalized Decision Matrix Data						
	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$
	Yield of extraction	A <sub>260/280</sub> ratio	A <sub>260/230</sub> ratio	Degree of DNA degradation	Easiness of amplification	Duration of extraction	Cost per extraction
UltraClean Conventional	0.20 <sup>‡</sup>	0.37	0.42	0.30	0.28	0.16	0.31
UltraClean Alternative	0.14	0.34	0.39	0.15	0.28	0.18	0.31
PowerSoil Conventional	0.16	0.35	0.41	0.15	0.28	0.19	0.37
PowerSoil Alternative	0.09	0.37	0.66	0.15	0.28	0.21	0.37
FastSpin Conventional	0.43	0.36	0.08	0.46	0.42	0.22	0.43
FastSpin Alternative	0.56	0.36	0.11	0.46	0.42	0.24	0.43
E.Z.N.A. Conventional	0.10	0.35	0.14	0.46	0.42	0.61	0.29
E.Z.N.A. Alternative	0.63	0.33	0.17	0.46	0.42	0.62	0.29

<sup>†</sup>UltraClean: UltraClean™ Soil DNA Isolation Kit; PowerSoil: PowerSoil® DNA Isolation Kit; FastSPIN: FastDNA® SPIN Kit for Soil; E.Z.N.A.: E.Z.N.A.® Soil DNA Isolation Kit. <sup>‡</sup>Example of calculation: Yield of extraction criterion for soil (using **Table 3**):  $0.20 = 8.21/[(8.21)^2 + (5.59)^2 + (6.76)^2 + (3.78)^2 + (17.70)^2 + (23.29)^2 + (4.33)^2 + (25.98)^2 + (5.24)^2 + (5.49)^2]^{0.5}$ .

**Table 9.** Summary of the weighted normalized decision matrix data for the TOPSIS method using seven decision criteria for each DNA extraction method for soil.

Methods <sup>†</sup>	Vij						
	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>	C <sub>5</sub>	C <sub>6</sub>	C <sub>7</sub>
	Yield of extraction	A <sub>260/280</sub> ratio	A <sub>260/230</sub> ratio	Degree of DNA degradation	Easiness of amplification	Duration of extraction	Cost per extraction
UltraClean Conventional	0.06 <sup>‡</sup>	0.00	0.12	0.04	0.01	0.03	0.01
UltraClean Alternative	0.04	0.00	0.11	0.02	0.01	0.04	0.01
PowerSoil Conventional	0.05	0.00	0.11	0.02	0.01	0.04	0.01
PowerSoil Alternative	0.03	0.00	0.19	0.02	0.01	0.04	0.01
FastSpin Conventional	0.14	0.00	0.02	0.06	0.01	0.05	0.01
FastSpin Alternative	0.18	0.00	0.03	0.06	0.01	0.05	0.01
E.Z.N.A. Conventional	0.03	0.00	0.04	0.06	0.01	0.13	0.01
E.Z.N.A. Alternative	0.20	0.00	0.05	0.06	0.01	0.13	0.01
Sum	0.74	0.00	0.67	0.36	0.08	0.51	0.06
V+ <sup>§</sup>	0.20	0.00	0.19	0.06	0.01	0.13	0.01
V- <sup>¶</sup>	0.03	0.00	0.02	0.02	0.01	0.03	0.01

<sup>†</sup>UltraClean: UltraClean™ Soil DNA Isolation Kit; PowerSoil: PowerSoil® DNA Isolation Kit; FastSPIN: FastDNA® SPIN Kit for Soil; E.Z.N.A.: E.Z.N.A.® Soil DNA Isolation Kit. <sup>‡</sup>Example of calculation:  $0.06 = 0.20$  (normalized data for UltraClean Conventional under the yield of extraction criterion, **Table 8**)  $\times$  0.32 (weight of yield of extraction criterion for soil, **Table 7**). <sup>§</sup>Ideal positive solution. <sup>¶</sup>Ideal negative solution.

**Table 10.** Summary of the positive and negative distances and the final TOPSIS scores of DNA extractions methods for soil.

Methods <sup>†</sup>	$d_i^+$	$d_i^-$	TOPSIS Score : $C_i + \text{Rank}$	
	Value	Value	Value	
UltraClean Conventional	0.156 <sup>‡</sup>	0.140 <sup>§</sup>	0.474 <sup>¶</sup>	4
UltraClean Alternative	0.175	0.134	0.433	6
PowerSoil Conventional	0.165	0.136	0.452	5
PowerSoil Alternative	0.172	0.189	0.522	1
FastSpin Conventional	0.181	0.136	0.429	7
FastSpin Alternative	0.164	0.171	0.511	2
E.Z.N.A. Conventional	0.247	0.018	0.068	8
E.Z.N.A. Alternative	0.176	0.174	0.496	3

<sup>†</sup>UltraClean: UltraClean™ Soil DNA Isolation Kit; PowerSoil: PowerSoil® DNA Isolation Kit; FastSPIN: FastDNA® SPIN Kit for Soil; E.Z.N.A.: E.Z.N.A.® Soil DNA Isolation Kit. Example of calculation <sup>‡</sup>using **Table 9**:  $0.156 = (0.06 - 0.20)^2 + (0.00 - 0.00)^2 + (0.12 - 0.19)^2 + (0.04 - 0.06)^2 + (0.01 - 0.01)^2 + (0.03 - 0.013)^2 + (0.01 - 0.01)^2$ ; <sup>§</sup> $0.140 = (0.06 - 0.03)^2 + (0.00 - 0.00)^2 + (0.12 - 0.02)^2 + (0.04 - 0.02)^2 + (0.01 - 0.01)^2 + (0.03 - 0.03)^2 + (0.01 - 0.01)^{0.5}$ ; <sup>¶</sup> $0.474 = 0.140 / (0.140 + 0.156)$ .

the application of an alternative lysis step for most of the extraction methods did not improve their performance significantly, except for the E.Z.N.A. method (**Table 3**). This method, both its conventional and alternative versions, resulted in high values of DNA purity. However, the E.Z.N.A. conventional method provided much lower yields of extraction in comparison to its alternative method, as was also observed in other investigations [32,33].

Because of the demonstrated conflicts among different

**Table 11.** Summary of the positive and negative distances and the final TOPSIS scores of DNA extractions methods for SM [soil:manure, 99:1(w/w)].

Methods <sup>†</sup>	$d_i^+$	$d_i^-$	TOPSIS Score : $C_i + \text{Rank}$	
	Value	Value	Value	
UltraClean Conventional	0.056	0.176	0.757	2
UltraClean Alternative	0.087	0.160	0.647	3
PowerSoil Conventional	0.055	0.184	0.770	1
PowerSoil Alternative	0.086	0.153	0.640	4
FastSpin Conventional	0.125	0.138	0.524	6
FastSpin Alternative	0.119	0.133	0.529	5
E.Z.N.A. Conventional	0.168	0.095	0.362	7
E.Z.N.A. Alternative	0.183	0.067	0.268	8

<sup>†</sup>UltraClean: UltraClean™ Soil DNA Isolation Kit; PowerSoil: PowerSoil® DNA Isolation Kit; FastSPIN: FastDNA® SPIN Kit for Soil; E.Z.N.A.: E.Z.N.A.® Soil DNA Isolation Kit.

extraction methods and criteria, the selection of the optimum method for different types of soil was not straightforward (**Table 4**). For this reason, for the first time in the field, the application of a systematic MCDM approach was proposed and implemented to select overall the optimum DNA extraction method for each type of soil. The PowerSoil method was systematically defined as the best performing method for extracting DNA from soil samples; more specifically, for soil, the alternative version of the PowerSoil method gained the first rank (**Table 10**), while for SM and SMB its conventional ver-

**Table 12.** Summary of the positive and negative distances and the final TOPSIS scores of DNA extraction methods for SMB [soil:manure:biochar, 98:1:1(w/w)].

Methods <sup>†</sup>	$d_i^+$	$d_i^-$	TOPSIS Score : $C_i$ + Rank	
	Value	Value	Value	
UltraClean Conventional	0.065	0.162	0.714	2
UltraClean Alternative	0.080	0.160	0.666	4
PowerSoil Conventional	0.059	0.172	0.744	1
PowerSoil Alternative	0.072	0.153	0.681	3
FastSpin Conventional	0.108	0.141	0.567	5
FastSpin Alternative	0.116	0.132	0.531	6
E.Z.N.A. Conventional	0.176	0.088	0.333	7
E.Z.N.A. Alternative	0.193	0.030	0.134	8

<sup>†</sup>UltraClean: UltraClean™ Soil DNA Isolation Kit; PowerSoil: PowerSoil® DNA Isolation Kit; FastSPIN: FastDNA® SPIN Kit for Soil; E.Z.N.A.: E.Z.N.A.® Soil DNA Isolation Kit.

sion performed best (Tables 11 and 12, respectively).

The performances of the DNA extraction kits assessed in the present study (Table 3) were similar or better than those reported in the literature. For the PowerSoil kit, which ranked first in the present work, the yields of extraction ranged between 3.78 and 11.97 µg of DNA/g of soil, while the  $A_{260/280}$  and  $A_{260/230}$  ratios varied from 1.55 to 1.96 and 0.86 to 2.07, respectively. Previous studies reported yields of extraction of 0.12 to 23.0 µg of DNA/g of soil, as well as  $A_{260/280}$  and  $A_{260/230}$  ratios of 1.34 - 1.65 and 0.55 - 0.61, respectively [10,14,32,33].

Using the UltraClean kit, we obtained yields of extraction ranging from 5.59 to 14.21 µg of DNA/g of soil,  $A_{260/280}$  ratios between 1.50 and 1.77, and  $A_{260/230}$  ratios between 0.81 and 1.43. Other authors presented yields of extraction between 0.31 and 2.81 µg of DNA/g of soil, an average  $A_{260/280}$  ratio of 1.33, and  $A_{260/230}$  ratios between 0.67 and 2.20 [10,29,31,34].

In the present study, the yields of extraction with the FastSpin kit ranged between 17.70 and 23.87 µg of DNA/g of soil, whereas the  $A_{260/280}$  and  $A_{260/230}$  ratios were 1.70 - 1.84 and 0.25 - 0.60, respectively. The literature indicates lower yields of extraction (0.80 to 9.12 µg of DNA/g of soil),  $A_{260/280}$  ratios (1.53 to 1.64) and  $A_{260/230}$  ratios (0.24 to 0.28) [10,29-31,33].

Regarding the E.Z.N.A. kit, Table 3 indicates yields of extraction of 4.33 to 25.98 µg of DNA/g of soil (with significant improvements using the alternative method),  $A_{260/280}$  ratios of 1.56 to 1.87, and  $A_{260/230}$  ratios of 0.30 to 1.70. Previous publications included similar or lower yields of extraction (between 0.60 to 12.5 µg of DNA/g of soil), similar  $A_{260/280}$  ratios ranging from 1.75 to 1.87, and generally higher  $A_{260/230}$  ratios varying from 1.59 to 1.87 [32,33].

## 5. CONCLUDING REMARK

In summary, the choice of a DNA extraction method for microbial ecology studies is critical to obtain reliable results since each method can affect the composition and the richness of microbial communities of tested samples. Hence, in selecting a suitable extraction method, it is necessary to take into account the type of the environmental sample (in our case, soil, SM, and SMB), quantitative and qualitative characteristics of extracted DNA (e.g., yield of extraction, purity, degradation degree, quality of PCR products), and downstream molecular analyses such as competitive PCR, real-time PCR, denaturing gradient gel electrophoresis (DGGE) and large-scale parallel-pyrosequencing. Based on the results of the performed case study, overall we recommend the PowerSoil® DNA Isolation Kit as an optimum method for obtaining total bacterial DNA from soil and soil-containing mixtures such as soil:manure and soil:manure:biochar. The standardization/selection of DNA extraction techniques in the field is a current problem, and hence the powerful MCDM approaches such as entropy/TOPSIS which were used in this study are recommended as a first step towards comparing similar methods in other studies.

## 6. ACKNOWLEDGEMENTS

This work was supported by the National Sciences and Engineering Research Council of Canada (NSERC—Discovery Grant) and the Fonds de recherche du Quebec sur la Nature et les Technologies (FRQNT—Team Research Grant Project) for operating funds, and by the Canada Foundation for Innovation (CFI—Leaders Opportunity Fund) for infrastructure funds to Martin R. Chénier. Snizhana V. Olishchevska benefited from the CFI—Infrastructure Operating Fund for personal support. The financial support by UBC's Work-Study Program to Sepideh Pakpour is also greatly acknowledged.

## REFERENCES

- [1] Robe, P., Nalin, R., Capellano, C., Vogel, T.A. and Simonet, P. (2003) Extraction of DNA from soil. *European Journal of Soil Biology*, **39**, 183-190. [http://dx.doi.org/10.1016/S1164-5563\(03\)00033-5](http://dx.doi.org/10.1016/S1164-5563(03)00033-5)
- [2] Verheijen, F.G.A., Jeffery, S., Bastos, A.C., van der Velde, M. and Diafas, I. (2009) Biochar application to soils - a critical scientific review of effects on soil properties, processes and functions. EUR 24099 EN, Office for the Official Publications of the European Communities, Luxembourg.
- [3] Amann, R.I., Ludwig, W. and Schleifer, K.H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiology Reviews*, **59**, 143-169.
- [4] Roose-Amsaleg, C.L., Garnier-Sillam, E. and Harry, M. (2001) Extraction and purification of microbial DNA from soil and sediment samples. *Applied Soil Ecology*, **18**,

- 47-60. [http://dx.doi.org/10.1016/S0929-1393\(01\)00149-4](http://dx.doi.org/10.1016/S0929-1393(01)00149-4)
- [5] Torsvik, V., Goksoyr, J. and Daae, F.L. (1990) High diversity in DNA of soil bacteria. *Applied and Environmental Microbiology*, **56**, 782-787.
- [6] Ritz, K. (2007) The plate debate: Cultivable communities have no utility in contemporary environmental microbial ecology. *Fems Microbiology Ecology*, **60**, 358-362. <http://dx.doi.org/10.1111/j.1574-6941.2007.00331.x>
- [7] Morris, C.E., Bardin, M., Berge, O., Frey-klett, P., Fromin, N., Girardin, H., Guinebretiére, M.H., Lebaron, P., Thiéry, J.M. and Trousselier, M. (2002) Microbial biodiversity: Approaches to experimental design and hypothesis testing in primary scientific literature from 1975 to 1999. *Microbiology and Molecular Biology Reviews*, **66**, 592-616. <http://dx.doi.org/10.1128/MMBR.66.4.592-616.2002>
- [8] Dorigo, U., Volatier, L. and Humbert, J.F. (2005) Molecular approaches to the assessment of biodiversity in aquatic microbial communities. *Water Research*, **39**, 2207-2218. <http://dx.doi.org/10.1016/j.watres.2005.04.007>
- [9] De Liphay, J.R., Enzinger, C., Johnsen, K., Aamand, J., and Sorensen, S.J. (2004) Impact of DNA extraction method on bacterial community composition measured by denaturing gradient gel electrophoresis. *Soil Biology & Biochemistry*, **36**, 1607-1614. <http://dx.doi.org/10.1016/j.soilbio.2004.03.011>
- [10] Ning, J., Liebich, J., Kastner, M., Zhou, J.Z., Schaffer, A. and Burauel, P. (2009) Different influences of DNA purity indices and quantity on PCR-based DGGE and functional gene microarray in soil microbial community study. *Applied Microbiology and Biotechnology*, **82**, 983-993. <http://dx.doi.org/10.1007/s00253-009-1912-0>
- [11] Carrigg, C., Rice, O., Kavanagh, S., Collins, G. and O'Flaherty, V. (2007) DNA extraction method affects microbial community profiles from soils and sediment. *Applied Microbiology and Biotechnology*, **77**, 955-964. <http://dx.doi.org/10.1007/s00253-007-1219-y>
- [12] Roh, C., Villatte, F., Kim, B.G. and Schmid, R.D. (2006) Comparative study of methods for extraction and purification of environmental DNA from soil and sludge samples. *Applied Biochemistry and Biotechnology*, **134**, 97-112. <http://dx.doi.org/10.1385/ABAB:134:2:97>
- [13] Arbeli, Z. and Fuentes, C.L. (2007) Improved purification and PCR amplification of DNA from environmental samples. *Fems Microbiology Letters*, **272**, 269-275. <http://dx.doi.org/10.1111/j.1574-6968.2007.00764.x>
- [14] Rosa, M.M., Tornisielo, S.M.T. and Ceccato-Antonini, R. (2010) Evaluation of different techniques for DNA direct extraction from Brazilian agricultural soil. *HOLOS Environment*, **10**, 12-25.
- [15] Burgmann, H., Pesaro, M., Widmer, F. and Zeyer, J. (2001) A strategy for optimizing quality and quantity of DNA extracted from soil. *Journal of Microbiological Methods*, **45**, 7-20. <http://dx.doi.org/10.1007/s11104-007-9193-9>
- [16] Steiner, C., Teixeira, W.G., Lehmann, J., Nehls, T., de Macedo, J.L.V., Blum, W.E.H. and Zech, W. (2007) Long term effects of manure, charcoal and mineral fertilization on crop production and fertility on a highly weathered Central Amazonian upland soil. *Plant and Soil*, **291**, 275-290. <http://dx.doi.org/10.1007/s11104-007-9193-9>
- [17] Mathews, J.A. (2008) Carbon-negative biofuels. *Energy Policy*, **36**, 940-945. <http://dx.doi.org/10.1016/j.enpol.2007.11.029>
- [18] Laird, D.A. (2008) The charcoal vision: A win-win-win scenario for simultaneously producing bioenergy, permanently sequestering carbon, while improving soil and water quality. *Agronomy Journal*, **100**, 178-181. <http://dx.doi.org/10.2134/agrojn12007.0161>
- [19] Sohi, S., Loez-Capel, E., Krull, E. and Bol, R. (2009) Biochar's roles in soil and climate change: A review of research needs. Okehampton, UK.
- [20] Tebbe, C.C. and Vahjen, W. (1993) Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant-DNA from bacteria and a yeast. *Applied and Environmental Microbiology*, **59**, 2657-2665.
- [21] Sørensen, S.J., Müller, A.K., Hansen, L.H., Rasmussen, L.D., Liphay, J.R. and Barkay, T. (2002) Molecular methods for assessing and manipulating the diversity of microbiological populations and processes. In: Burns RG, Dick RP, Eds., *Enzymes in the Environment*, Marcel Dekker, New York. <http://dx.doi.org/10.1201/9780203904039.ch14>
- [22] Fortin, N., Beaumier, D., Lee, K. and Greer, C.W. (2004) Soil washing improves the recovery of total community DNA from polluted and high organic content sediments. *Journal of Microbiological Methods*, **56**, 181-191. <http://dx.doi.org/10.1016/j.mimet.2003.10.006>
- [23] Wu, L.Y., Thompson, D.K., Li, G.S., Hurt, R.A., Tiedje, J.M. and Zhou, J. (2001) Development and evaluation of functional gene arrays for detection of selected genes in the environmental. *Applied and Environmental Microbiology*, **67**, 5780-5790. <http://dx.doi.org/10.1128/AEM.67.12.5780-5790.2001>
- [24] Zhou, J.Z. and Thompson, D.K. (2002) Challenges in applying microarrays to environmental studies. *Current Opinion in Biotechnology*, **13**, 204-207. [http://dx.doi.org/10.1016/S0958-1669\(02\)00319-1](http://dx.doi.org/10.1016/S0958-1669(02)00319-1)
- [25] Holben, W.E., Jansson, J.K., Chelm, B.K. and Tiedje, J.M. (1988) DNA probe method for the detection of specific microorganisms in the soil bacterial community. *Applied and Environmental Microbiology*, **54**, 703-711.
- [26] Jacobsen, C.S. and Rasmussen, O.F. (1992) Development and application of a new method to extract bacterial-DNA from soil based on separation of bacteria from soil with cation-exchange resin. *Applied and Environmental Microbiology*, **58**, 2458-2462.
- [27] Yeates, C., Gillings, M.R., Davison, A.D., Altavilla, N. and Veal, D.A. (1998) Methods for microbial DNA extraction from soil for PCR amplification. *Biological Procedures Online*, **14**, 1-8. <http://dx.doi.org/10.1251/bpo6>
- [28] Kumar, P.P. (2011) An improved method for extraction of heterologous DNA from environmental samples for the construction of metagenomic libraries. *International Journal of Pharma and Bio Sciences*, **2**, 368-373.



- [29] Martin-Laurent, F., Philippot, L., Hallet, S., Chaussod, R., Germon, J.C., Soulas, G., and Catroux, G. (2001) DNA extraction from soils: Old bias for new microbial diversity analysis methods. *Applied and Environmental Microbiology*, **67**, 2354-2359. <http://dx.doi.org/10.1128/AEM.67.5.2354-2359.2001>
- [30] Park, J.W. and Crowley, D.E. (2005) Normalization of soil DNA extraction for accurate quantification real-time PCR and of target genes by DGGE. *Biotechniques*, **38**, 579-586. <http://dx.doi.org/10.2144/05384ST04>
- [31] Klerks, M.M., van Bruggen, A.H.C, Zijlstra, C. and Donnikov, M. (2006) Comparison of methods of extracting *Salmonella enterica* serovar enteritidis DNA from Model Based on Aras-G and AHP Methods for Multiple environmental substrates and quantification of organisms by using a general internal procedural control. *Applied and Environmental Microbiology*, **72**, 3879-3886. <http://dx.doi.org/10.1128/AEM.02266-05>
- [32] Huang, B., Zu, T. and Guo, Q. (2007) Extract PCR-ready soil DNA with the new E.Z.N.A.<sup>®</sup> soil DNA isolation kit. The Market Source for Life Science VWR biomarker International, 26-27.
- [33] Dineen, S.M., Aranda, R., Anders, D.L., and Robertson, J.M. (2010) An evaluation of commercial DNA extraction kits for the isolation of bacterial spore DNA from soil. *Journal of Applied Microbiology*, **109**, 1886-1896. <http://dx.doi.org/10.1111/j.1365-2672.2010.04816.x>
- [34] Fitzpatrick, K.A., Kersh, G.J. and Massung, R.F. (2010) Practical Method for Extraction of PCR-Quality DNA from Environmental Soil Samples. *Applied and Environmental Microbiology*, **76**, 4571-4573. <http://dx.doi.org/10.1128/AEM.02825-09>
- [35] Roy, B. (1993) Aide Multicritère à la Décision: Méthodes et Cas. Economica, Paris.
- [36] Roy, B. (1990) Decision-aid and decision-making. In: Bana, E. and Costa, C.A., Eds., *Readings in Multiple Criteria Decision-Aid*. Springer-Verlag, Heidelberg, 17-35. [http://dx.doi.org/10.1007/978-3-642-75935-2\\_2](http://dx.doi.org/10.1007/978-3-642-75935-2_2)
- [37] Collette, Y. and Siarry, P. (2003) Multiobjective optimization. Springer-Verlag, New York.
- [38] Pakpour, S., Milani, A. and Chénier, M. (2012) A multi-criteria decision-making approach for comparing sample preservation and DNA extraction methods from swine feces. *American Journal of Molecular Biology*, **2**, 159-169. <http://dx.doi.org/10.4236/ajmb.2012.22018>
- [39] Putrus, P. (1990) Accounting for intangibles in integrated manufacturing-nonfinancial justification based on the analytical hierarchy process. *Information Strategy*, **6**, 25-30.
- [40] Boucher, T.O. and McStravic, E.L. (1991) Multi-attribute evaluation within a present value framework and its relation to the analytic hierarchy process. *The Engineering Economist*, **37**, 55-71. <http://dx.doi.org/10.1080/00137919108903055>
- [41] Pohekar, S.D. and Ramachandran, M. (2004) Application of multi-criteria decision making to sustainable energy planning—A review. *Renewable and Sustainable Energy Reviews*, **8**, 365-381. <http://dx.doi.org/10.1016/j.rser.2003.12.007>
- [42] Turskis, Z., Zavadskas, K.E. and Kutut, V. (2013) A criteria prioritizing of heritage value. *International Journal of Information Technology and Decision Making*, **12**, 45. <http://dx.doi.org/10.1142/S021962201350003X>
- [43] Hwang, C.L. and Yoon, K. (1981) Multiple attribute decision making: Methods and applications. Springer-Verlag, New York. <http://dx.doi.org/10.1007/978-3-642-48318-9>
- [44] Ewa, R. (2011) Multi-criteria decision making models by applying the TOPSIS method to crisp and interval data. In: Trzaskalik, T. and Wachowicz, T., Eds., *Multiple Criteria Decision Making*, The University of Economics, Katowice, pp. 200-230.
- [45] Muyzer, G., Dewaal, E.C. and Uitterlinden, A.G. (1993) Profiling of complex microbial-populations by denaturing gradient gel-electrophoresis analysis of polymerase chain reaction-amplified genes-coding for 16s ribosomal-RNA. *Applied and Environmental Microbiology*, **59**, 695-700.
- [46] Chenier, M.R. and Juteau, P. (2009) Fate of chlortetracycline- and tylosin-resistant bacteria in an aerobic thermophilic sequencing batch reactor treating swine waste. *Microbial Ecology*, **58**, 86-97. <http://dx.doi.org/10.1007/s00248-008-9478-4>
- [47] Thermo Fisher Scientific (2009) NanoDrop 2000/2000c spectrophotometer, V1.0 user manual.
- [48] Ahn, B.S. (2011) Compatible weighting method with rank order centroid: Maximum entropy ordered weighted averaging approach. *European Journal of Operational Research*, **212**, 552-559. <http://dx.doi.org/10.1016/j.ejor.2011.02.017>
- [49] Zou, Z.H., Yan, Y. and Sun, J.N. (2006) Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation for water quality assessment. *Journal of Environmental Sciences-China*, **18**, 1020-1023. [http://dx.doi.org/10.1016/S1001-0742\(06\)60032-6](http://dx.doi.org/10.1016/S1001-0742(06)60032-6)
- [50] Lemarchand, K., Berthiaume, F., Maynard, C., Harel, J., Payment, P., Bayardelle, P., Masson, L. and Brousseau, R. (2005) Optimization of microbial DNA extraction and purification from raw wastewater samples for downstream pathogen detection by microarrays. *Journal of Microbiological Methods*, **63**, 115-126. <http://dx.doi.org/10.1016/j.mimet.2005.02.021>



# Simulating Dryland Water Availability and Spring Wheat Production in the Northern Great Plains

Zhiming Qi,\* Patricia N. S. Bartling, Jalal D. Jabro, Andrew W. Lenssen, William M. Iversen, Lajpat R. Ahuja, Liwang Ma, Brett L. Allen, and Robert G. Evans

## ABSTRACT

Agricultural system models are useful tools to synthesize field experimental data and to extrapolate results to longer periods of weather and to other cropping systems. The objectives of this study were: (i) to quantify the effects of crop management practices and tillage on soil water and spring wheat (*Triticum aestivum* L.) production in a continuous spring wheat system using the RZWQM2 model (coupled with CERES-Wheat) under a dryland condition, and (ii) to extend the RZWQM2 model results to longer term weather conditions and propose alternate cropping systems and management practices. Measured soil water content, yield, and total aboveground biomass under different tillage and plant management practices were used to calibrate and evaluate the RZWQM2 model. The model showed no impacts of tillage but late planting greatly reduced grain yield and biomass, in agreement with observed differences among treatments. The hydrologic analysis under long-term climate variability showed a large water deficit (32.3 cm) for spring wheat. Fallowing the cropland every other year conserved 4.2 cm of water for the following wheat year, of which only 1.7 cm water was taken up by wheat, resulting in a yield increase of 249 kg ha<sup>-1</sup> (13.7%); however, the annualized mean yield decreased 782 kg ha<sup>-1</sup> (43.1%) due to 1 yr of fallow. Other long-term simulations showed that optimal planting dates ranged from 1 March to 10 April and the seeding rates with optimum economic return were 3.71 and 3.95 × 10<sup>6</sup> seeds ha<sup>-1</sup> for conventional and ecological management treatments, respectively.

**S**PRING WHEAT (*Triticum aestivum* L., excluding durum wheat) was harvested on 5.35 million ha of croplands in the United States in 2011, with 89.1% occurring in the northern Great Plains (NGP) states including North Dakota, Montana, Minnesota, and South Dakota. Montana had the highest percentage of spring wheat acreage (31.3%) among all spring wheat production states (National Agricultural Statistics Service, 2011a). The 2007 Census of Agriculture showed that about half of the spring wheat acreage in Montana was under a continuous spring wheat production system, while the other half was in a 2-yr spring wheat–fallow system. Most of this spring wheat area (95.4%) was rainfed (National Agricultural Statistics Service, 2007).

The lack of available water for crop growth is the primary factor affecting dryland spring wheat production in the NGP. A spring wheat yield trial conducted at Sidney, MT, with >70 cultivars during 2005 to 2009 showed that the yield of dryland wheat was about 30% less than that of the irrigated crop (J. Eckoff, personal communication, 2010). Brown et al. (1981) reported that spring wheat yield increased 135 kg ha<sup>-1</sup> with every centimeter increase

in plant water use in Montana and North Dakota. A similar relationship between spring wheat yield and plant-available water was also found in the inland Pacific Northwest (Schillinger et al., 2008). Winter wheat yields in the central Great Plains increased by 141 kg ha<sup>-1</sup> for every centimeter increase in plant-available water in the soil at planting (Nielsen et al., 2002) and by 125 kg ha<sup>-1</sup> for every centimeter of water uptake after 13 cm of water use (Nielsen et al., 2011).

Various management strategies have been proposed and applied to cope with soil water shortage for dryland spring wheat production in the NGP, including no-till and reduced tillage with residue mulching and crop rotations. Fenster (1973) reported that the soil water storage efficiency increased from 16 to 31% in Montana by adding surface residue cover during summer fallow. Nielsen and Vigil (2010) reported that precipitation storage efficiency during the fallow period of a winter wheat–fallow system increased from 20% with conventional tillage fallow management to 35% for no-till management in Colorado. In general, a higher percentage of residue cover would lead to higher soil water storage (Tanaka and Aase, 1987). Lenssen et al. (2007) documented that zero tillage often provided higher soil water content at planting; however, Deibert et al. (1986) found that a difference in water storage between no-till and tilled field in North Dakota was not evident

Z. Qi, P.N.S. Bartling, L.R. Ahuja, L. Ma, USDA-ARS Agricultural Systems Research Unit, Fort Collins, CO 80526; Z. Qi currently at Dep. of Bioresource Engineering, Macdonald Campus, McGill Univ., Ste-Anne-de-Bellevue, QC H9X 3V9, Canada; J.D. Jabro, W.M. Iversen, B.L. Allen, R.G. Evans, USDA-ARS Northern Plains Agricultural Research Lab., Sidney, MT 59270; and A.W. Lenssen, Dep. of Agronomy, Iowa State Univ., Ames, IA 50011. Received 4 June 2012. \*Corresponding author (zhiming.qi@mcgill.ca).

Published in *Agron. J.* 105:37–50 (2013)  
doi:10.2134/agronj2012.0203

Copyright © 2013 by the American Society of Agronomy, 5585 Guilford Road, Madison, WI 53711. All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher.

**Abbreviations:** CERES, Crop Estimation through Resource and Environment Synthesis; CTC, conventional tillage, conventional management; CTE, conventional tillage, ecological management; DAP, days after planting; DSSAT, Decision Support System for Agrotechnology Transfer; ET, evapotranspiration; NGP, northern Great Plains; NSE, Nash–Sutcliffe model efficiency; NTC, no-till, conventional management; NTE, no-till, ecological management; PBIAS, percent bias; RMSD, root mean squared deviation; RS<sub>p</sub>R, root mean squared deviation/pooled sample variance ratio; RZWQM2, Root Zone Water Quality Model 2.

and neither was a difference in evapotranspiration between continuous spring wheat and spring wheat–fallow treatments.

For better weed control, a delay in planting from mid-April to mid-May for spring wheat has been proposed to allow mechanical weed control before planting (Sainju et al., 2011). This delay, however, reduces the length of the growth period, which could reduce biomass accumulation. Furthermore, the associated changes in phenology might alter the timing of water stress relative to plant development, possibly reducing dryland crop yield. To compensate for a short growing season and less biomass accumulation, and to suppress weed competition, a higher seeding rate is used when the spring wheat is planted in mid-May. The late planting date along with a high seeding rate is called *ecological management*, as opposed to the conventional management with a customary planting date and seeding rate.

Agricultural system models, which include the interactions among the various processes and factors in the system, are useful tools to evaluate various agronomic management practices after careful calibration. The Root Zone Water Quality Model 2 (RZWQM2) including Decision Support System for Agrotechnology Transfer (DSSAT) version 4.0 cropping system models has been successfully used to simulate water availability and crop production under long-term weather conditions and various management practices. Thorp et al. (2007) and Qi et al. (2011) documented that the RZWQM2 model adequately simulated hydrology, crop yield, and N dynamics during the evaluation period. The model was then used to predict crop production and water balance during 40 to 45 yr. Using the successfully validated RZWQM2 model, Ma et al. (2007) and Malone et al. (2007) simulated crop yield and water quality under additional fertilization, drainage, and crop management practices. In the semiarid Great Plains, Saseendran et al. (2005) reported that both RZWQM2 and Crop Estimation through Resource and Environment Synthesis (CERES)-Maize models accurately predicted the observed decline in corn (*Zea mays* L.) yield with delayed planting dates. The RZWQM2 model adequately simulated the yield and biomass for dryland winter wheat, corn, and proso millet (*Panicum miliaceum* L. ssp. *miliaceum*) in semiarid eastern Colorado, and the model was subsequently used to simulate crop yield under potential crop rotations (Saseendran et al., 2010).

Limited information is available on the simulation of spring wheat water use and yield in the semiarid NGP. Chipanshi et al. (1997) reported a successful simulation study for spring wheat production using CERES-Wheat version 3.5 at three sites in central and southern Saskatchewan, Canada, and the ratio of simulated to observed total aboveground biomass at various growth stages ranged from 0.64 to 1.62 across the three locations. An earlier modeling study conducted in Saskatchewan suggested, however, that both CERES and Erosion/Productivity Impact Calculator (EPIC) simulated the annual spring wheat yield poorly, even though the predicted long-term mean yield was reasonable (Moulin and Beckie, 1993). The purpose of this study was to simulate water availability and spring wheat production under conventional and ecological management practices in the NGP under both tilled and no-till treatments using RZWQM2 with DSSAT

version 4.0, CERES-Wheat model included. The specific objectives of this modeling study were to: (i) quantify the effect of planting date, seeding rate, and tillage on dryland spring wheat production in terms of soil water, yield, and biomass; and (ii) extend the results to longer term weather conditions and propose alternate cropping system and management practices.

## MATERIALS AND METHODS

### Root Zone Water Quality Model 2

The RZWQM2 (version 2.40) is a one-dimensional agricultural system model including hydrology, nutrient and pesticide transport and transformation, plant growth, and management practice components (Ahuja et al., 2000). Infiltration from rainfall, irrigation, and snowmelt is computed using the Green–Ampt equation. Water redistribution in the soil profile, considering plant uptake as a sink, is simulated by the Richards equation. When the incoming water flux exceeds the soil infiltration capacity, the difference will be diverted into macropore flow if there are macropores; otherwise this excessive water becomes off-site runoff. The nutrient chemistry processes model incorporated in RZWQM2 is OMNI (Shaffer et al., 2000). The DSSAT family (version 4.0) of cropping system models (CROPGRO and CERES) was incorporated into RZWQM2 (Ma et al., 2005, 2006), and the CERES-Wheat model (Jones et al., 2003; Ritchie et al., 1998) was used in this study.

### Field Experiment and Measurements

The field experiment was conducted from 2004 to 2010 on the Rasmussen dryland farm site, located 11 km west of Sidney (47°46' N, 104°16' W) in eastern Montana. The soil was mapped as a Williams loam (a fine-loamy, mixed, superactive, frigid Typic Argiustoll) formed in calcareous glacial till parent material with 0 to 4% slope. Wheat yield, aboveground biomass, and soil water content in continuous spring wheat treatments were compared under various management practices including tillage, planting date, seeding rate, and stubble height. The treatment factors were tillage (conventional tillage and no-till) and plant management (conventional and ecological management). In the ecological management, spring wheat was planted later but with a higher seeding rate compared with conventional management. This 2 × 2 factorial experiment design resulted in four treatments: conventional tillage, conventional management (CTC), conventional tillage, ecological management (CTE), no-till, conventional management (NTC), and no-till, ecological management (NTE). The treatments were arranged in three randomized complete blocks using a split-plot design with tillage practice as the main plots. There were three replications for each treatment, and each individual plot measured 12 by 12 m.

Conventional tillage before planting was one-pass field cultivation to a depth of 7 to 8 cm with C-shank sweeps at 45-cm spacing and 60-cm-length coil-tooth spring harrows. The conventional management treatment had customary planting dates (mid- to late April), with a customary seeding rate ( $2.22 \times 10^6$  seeds ha<sup>-1</sup>), while the ecological management treatment included later planting dates (about 3–4 wk later than conventional management) with a higher seeding rate ( $2.97 \times 10^6$  seeds ha<sup>-1</sup>). Urea fertilizer was broadcast before

planting for the conventionally managed spring wheat, while urea for the ecological management treatments was banded at planting 5 cm beneath and to the side of seed. Nitrogen fertilizer was applied at a rate of 107 kg N ha<sup>-1</sup> in 2004 to 2008 and at approximately 50 kg N ha<sup>-1</sup> in 2009 and 2010 based on a soil NO<sub>3</sub> test and target yield. Phosphate (as monoammonium phosphate at 11–52–0 [N–P<sub>2</sub>O<sub>5</sub>–K<sub>2</sub>O]) and potash (as KCl at 0–0–60 [N–P<sub>2</sub>O<sub>5</sub>–K<sub>2</sub>O]) were also banded at planting at 56 and 45 kg ha<sup>-1</sup>, respectively.

The climate data needed to run the RZWQM2 model were collected from an automated weather station starting in 2000 and are available at the Northern Plains Agricultural Research Laboratory weather network (<http://216.228.51.248/awn/>). Soil particle size distribution and bulk density were measured in 2010. Soil cores were used to measure the bulk density and water content from oven-dried undisturbed soil cores as the mass of oven-dried soil per volume of core. The particle size distribution was determined using the hydrometer method (Bouyoucos, 1962).

The soil water content was measured by lowering a neutron probe to five depths of 23, 46, 61, 91, and 122 cm. These measured soil moisture at these depths were used to represent the soil water content in soil layers of 0 to 34, 34 to 53, 53 to 76, 76 to 107, and 107 to 120 cm. The depth increment was shorter for the 0- to 76-cm depth to get a better resolution of measured soil moisture. The neutron probe was calibrated in the laboratory using a barrel method with soil obtained from the field site packed to the original bulk density (Chanasyk and Naeth, 1996). Total aboveground biomass was measured at harvest by clipping all the wheat plants within two randomly located 0.5-m<sup>2</sup> quadrats in each plot. For each quadrat, harvest tillers were counted and wheat grain was sampled to measure the kernel weight. A small-plot combine was used to harvest seven rows of wheat in the center of each plot to measure the grain yield.

### Model Inputs and Calibration

Field management parameters for the model simulation, such as sowing and harvest dates, seeding rates, fertilization, stubble height, as well as tillage, were obtained through field observation and management records. The weather data, including hourly precipitation, daily maximum and minimum air temperatures, solar radiation, wind speed, and relative humidity, were examined for outliers before being input into the model. Soil bulk density and soil texture were measured

from field samples obtained from the experimental site, while the hydraulic parameters in Brooks and Corey (1964) were determined from the observed soil water data. Crop growth and development parameters were calibrated against the observed yield and biomass data.

The calibration was conducted following the protocol provided by Ma et al. (2011, 2012). The protocol suggested two options for a model calibration strategy: select one treatment in one or multiple years or use multiple treatments in 1 yr. In this study, we chose to use the data from one treatment for all the years (2004–2010), which included wet (2010), average (2005), and dry (2008) years. The CTC treatment was selected as the calibration treatment. We first calibrated the model for soil moisture, then biomass, and finally yield. This parameterization sequence was then iterated three to four times. Besides comparing the simulated yield, biomass, and soil water with the measured values, we also checked model outputs for unmeasured variables such as anthesis and maturity dates to ensure that they were close to reported dates in the literature.

### Soil Hydraulic Parameters

Soil hydraulic parameters, which were mainly calibrated against soil water content data, as well as measured bulk density and soil texture, are listed in Table 1. The saturated hydraulic conductivity and residual soil water content were computed using the pedotransfer function package Rosetta using bulk density and soil texture (Schaap et al., 1998). The bubbling pressure (also known as the air-entry pressure,  $h_b$ ) and pore size distribution index ( $\lambda$ ) of the Brooks–Corey equation were initially taken from default values in Ma et al. (2011) and subsequently calibrated with the measured soil water content of each soil layer. In our case, the default  $h_b$  and  $\lambda$ , in general, led to an overestimation of soil moisture for all layers. The Brooks and Corey parameters for all soil layers were adjusted manually layer by layer from top to bottom. Our experience showed that  $\lambda$  was more sensitive than  $h_b$  in soil water simulation. The  $\lambda$  value was increased to reduce the overall simulated soil water content, and the  $h_b$  value was further reduced to get a better fit of low soil moisture during the summer.

### Crop Parameters

Crop parameters were manually adjusted to fit the measured biomass and yield components. Because phenology was not recorded in this study, parameters that affect the growing

**Table 1. Measured soil bulk density (BD) and texture and calibrated hydraulic properties of bubbling pressure ( $h_b$ ), pore size distribution index ( $\lambda$ ), saturated hydraulic conductivity ( $K_{sat}$ ), residual ( $\theta_r$ ) and saturated ( $\theta_s$ ) water contents, and water contents at 10 ( $\theta_{10}$ ), 33 ( $\theta_{33}$ ) and 1500 ( $\theta_{1500}$ ) kPa for Williams loam soil near Sidney, MT.**

Depth cm	BD g cm <sup>-3</sup>	Sand %	Silt %	Clay %	$h_b$ cm	$\lambda$	$K_{sat}$ cm h <sup>-1</sup>	Soil water content				
								$\theta_r$	$\theta_s$	$\theta_{10}$	$\theta_{33}$	$\theta_{1500}$
0–12	1.36	37	31	32	2.5	0.22	0.525	0.075	0.487	0.258	0.215	0.136
12–34	1.37	33	28	39	7.5	0.20	0.550	0.075	0.483	0.321	0.269	0.167
34–53	1.39	30	26	44	8.1	0.20	0.549	0.090	0.475	0.323	0.273	0.175
53–76	1.47	32	26	42	10	0.19	0.550	0.090	0.445	0.319	0.272	0.178
76–107	1.55	34	26	40	25	0.15	0.174	0.090	0.396	0.339	0.298	0.207
107–120	1.60	34	26	40	35	0.14	0.174	0.090	0.396	0.356	0.316	0.225
120–137	1.60	34	26	40	35	0.13	0.174	0.090	0.396	0.357	0.318	0.228
137–149	1.60	34	26	40	35	0.13	0.050	0.090	0.396	0.357	0.318	0.228



season length were estimated from the literature. The DSSAT 4.0 version of CERES-Wheat for spring wheat has not been tested in the NGP but an earlier version 3.5 was calibrated for southern Saskatchewan and Alberta, Canada (Chipanshi et al., 1997; Toure et al., 1995). The parameters in those studies with CERES version 3.5 were converted to equivalents for CERES version 4.0.

The parameter PIV (vernalization days) was set equal to zero because it is generally accepted that spring wheat does not need to be vernalized; in the literature, assigned model values of PIV ranged from 0 to 30 d (Toure et al., 1995; Chipanshi et al., 1997; McMaster et al., 2008). Although Sherman et al. (2010) stated that Reeder spring wheat was insensitive to photoperiod, the parameter PID (development reduction) was adjusted to a value of 21%, suggesting a low sensitivity to photoperiod, rather than zero, which would lead to a shorter simulated growing season length. This value is higher than 12% reported by Thorp et al. (2010). It is comparable to the 20% of Toure et al. (1995) for a long-term simulation in southern Alberta but much lower than the 60% suggested for spring wheat in the NGP (Godwin et al., 1990) and the calibrated value of 40% for spring wheat in Saskatchewan, Canada (Chipanshi et al., 1997). Because the anthesis or maturity dates were not recorded, we first used estimated phenology dates from the North Dakota Agricultural Weather Network (NDAWN) for spring wheat planted at Sidney, MT, during 2004 to 2010 and subsequently adjusted the thermal units (P1–P4) required in the vegetative stages, along with the grain-filling duration (P5), to get an average growing season length of 106 d after planting (DAP), which was reported by Talbert et al. (2001). Selecting the default ecotype of DS3585 in the CERES-Wheat ecotype database in general produced simulated phenology dates close to those estimated by NDAWN. The values for P1, P2, P3, P4, and P5 were finally set as 400, 350, 160, 300, and 400°C d, respectively, with some values comparable to results for winter wheat in Thorp et al. (2010). The thermal unit for the interval between successive leaf tip appearances (PHINT) was set as 100 to simulate seven or eight leaves at maturity.

After calibration of the crop development parameters, efforts were made to adjust other crop growth and yield parameters. The conversion rates of photosynthetically active radiation to dry matter at the vegetative and reproductive stages, PARUV and PARUR, respectively, were adjusted manually to obtain a good fit with the measured biomass and average measured harvest index of 0.36. Both parameters were set at 2.2 g MJ<sup>-1</sup> after several iterations with grain yield calibration because the biomass accumulated during vegetative growth and grain yield at maturity were interactively influenced by each other. The CERES-Wheat model predicts kernel number per head based on the stem weight at anthesis. In the crop yield component, the parameter G3, which controls the mortality or abortion rate of tillers experiencing water stress and heat stress, was set to 1.5 to better fit the observed average number of harvest tillers. The G1 and G2 parameters directly affect grain yield in the CERES-Wheat model: G1 controls the kernel number per unit canopy weight at anthesis and G2 is the standard kernel size under optimum conditions. Kernel weight at harvest was measured in this study, with values ranging from 12.7 mg kernel<sup>-1</sup> in a NTC plot in 2008 to 40.2 mg kernel<sup>-1</sup>

in another NTC plot in 2009. Theoretically, G2 should be set greater than the observed maximum value of 40.2 mg kernel<sup>-1</sup>. Attempts were made to simulate yield with an optimum kernel size value >40.2 mg kernel<sup>-1</sup>, but the kernel size was consistently overestimated for all years. To get a reasonable simulation in yield with a large kernel size, the kernel number (G1) had to be set equal to an extremely low value, which was out of the lower range of 15 to 30 kernels per unit gram weight of canopy in Ma et al. (2011). Therefore, in this study, we used the measured average kernel size of 30 mg kernel<sup>-1</sup> to get a reasonable simulation in kernel number and yield at harvest. The value of G1 was thereafter set at 15.8 kernels per unit of canopy weight for better grain yield simulation.

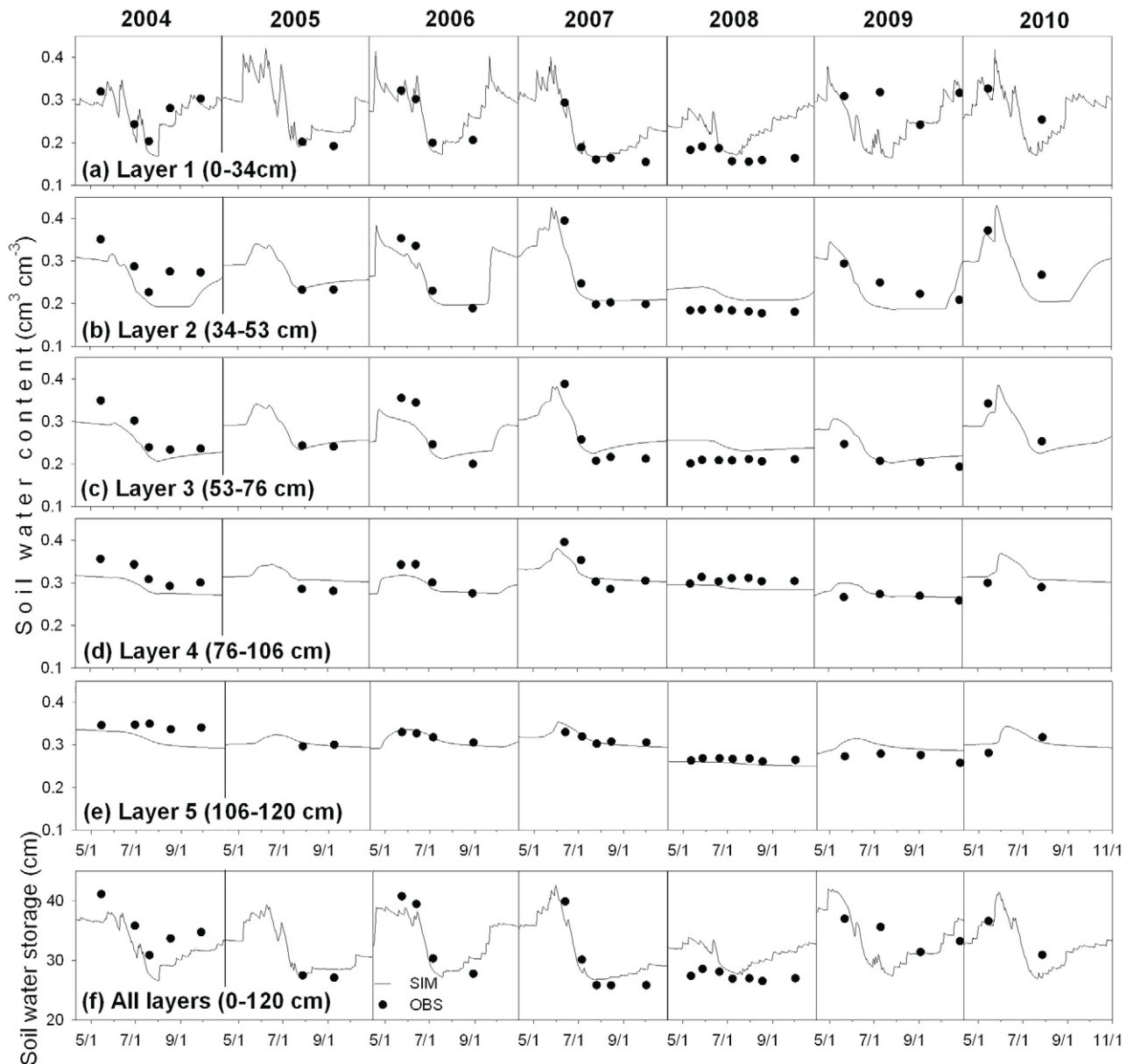
## Statistical Analysis

A number of statistical methods were used to quantify the goodness-of-fit of simulated data with observed information. In this study, we used percent bias (PBIAS), Nash–Sutcliffe model efficiency (NSE; Nash and Sutcliffe, 1970), root mean squared deviation (RMSD, also known as root mean squared error), and RMSD/pooled sample variance ratio (RS<sub>p</sub>R). Equations for these statistical approaches can be found in Ma et al. (2012).

Because one of the objectives of this study was to simulate the differences under various treatments, the model performance was considered acceptable when the simulation results reflected the measured treatment differences, as stated in Ma et al. (2011). For specific components of the simulation, such as yield, biomass, and soil water content, the criteria to justify the goodness-of-fit were set in accordance with the literature. In this study, model performance was defined “acceptable” when PBIAS was within ±15% (Ritchie et al., 1998; Hanson et al., 1999; Ahuja et al., 2000), NSE was >0.5 (Moriassi et al., 2007), and RS<sub>p</sub>R was ≤1.5. The RS<sub>p</sub>R limit indicates that the simulated error is less than 1.5 times the experiment error when RS<sub>p</sub>R is ≤1.5.

## Model Application to an Extended Climate

A 50-yr weather data set (1961–2010) for Sidney, MT, was used to assess the long-term impacts of current management practices (CTC, CTE, NTC, and NTE) on spring wheat yield production. The 50-yr weather data were collected from various sources. For the years from 2000 to 2010, all the weather information needed to drive RZWQM2 was site-specifically measured and downloaded from the Northern Plains Agricultural Research Laboratory website. For the years of 1961 to 1999, precipitation, air temperature, and solar radiation for Sidney, MT, were downloaded from the National Climate Data Center (NCDC; global and U.S. daily surface data, [www.ncdc.noaa.gov](http://www.ncdc.noaa.gov)) and the National Solar Radiation Database (NSRDB, [http://rredc.nrel.gov/solar/old\\_data/nsrdb/](http://rredc.nrel.gov/solar/old_data/nsrdb/)). Precipitation during 2000 to 2010 was recorded at 15-min intervals and aggregated to hourly precipitation, while during 1961 to 1999, precipitation was only available on a daily basis and was evenly distributed across 6 h in each day in which precipitation occurred. The wind speed and relative humidity for Sidney, MT, during 1961 to 1999 were obtained from the closest neighboring site at Williston, ND, which is 60 km from Sidney, MT, also available at the NCDC Climate Data



**Fig. 1.** Observed (OBS) and simulated (SIM) soil water content and soil water storage under spring wheat at Sidney, MT, for the calibration conventional tillage, conventional management treatment. Simulations were done with RZWQM2 with CERES-Wheat.

Online website. The relative humidity was calculated from dew point temperature and daily maximum and minimum air temperature according to Allen et al. (1998). Solar radiation data for Sidney, MT, was available only from 1991 on; for the years 1961 to 1990, the solar radiation at a neighboring site at Glasgow, MT, was used. Glasgow was the closest site that had the best quality of data (a Class I site as listed at the NSRDB website). We compared the solar radiation data for those two sites of Glasgow and Sidney. The 15-yr (1991–2005) average daily solar radiation was  $13.76 \text{ MJ m}^{-2} \text{ d}^{-1}$  for Sidney and  $13.78 \text{ MJ m}^{-2} \text{ d}^{-1}$  for Glasgow. For the spring wheat growing season (April–August), the average daily solar radiation was  $20.95$  and  $21.37 \text{ MJ m}^{-2} \text{ d}^{-1}$  for Sidney and Glasgow, respectively. Before being input into the model, quality control was done for all the data by plotting them vs. date to examine outliers and by calculating annual averages or sums to identify unusual trends.

We ran the calibrated model continuously with the long-term weather data to compare crop yield under continuous wheat–wheat and wheat–fallow cropping systems. For wheat–fallow, we simulated both wheat–fallow and fallow–wheat rotations and present the results as averages of the two rotations to eliminate weather differences between odd and even years. The calibrated model was also used to determine the optimal planting dates and an optimal seeding rate for spring wheat in Sidney, MT, by running the calibrated model for 50 yr (1961–2010) with planting dates varying from 20 February to 20 May each year with an interval of 10 or 11 d and a seeding rate ranging from  $1.73 \times 10^6$  to  $1.48 \times 10^7$  seeds  $\text{ha}^{-1}$ . Net return was calculated by subtracting seed cost from grain yield income for use in calculating the optimum seeding rate. We assumed a wheat grain price of  $\text{US}\$0.28 \text{ kg}^{-1}$  ( $\text{US}\$7.70 \text{ bu}^{-1}$ ) and a price for hard red spring wheat seed of  $\text{US}\$0.51 \text{ kg}^{-1}$  ( $\text{US}\$14.0 \text{ bu}^{-1}$ ) based on the local prices reported for March 2011 (National Agricultural Statistics Service, 2011b).



**Table 2. Statistics for comparison of modeled and observed soil water content and total soil water storage for the calibration (CTC) and evaluation (CTE, NTC, and NTE) treatments of spring wheat at Sidney, MT using RZWQM2 with CERES-Wheat.**

Treatment†	Statistics‡	Soil water content					Total soil water storage (0–120 cm)
		Layer 1 (0–34 cm)	Layer 2 (34–53 cm)	Layer 3 (53–76 cm)	Layer 4 (76–107 cm)	Layer 5 (107–120 cm)	
		cm <sup>3</sup> cm <sup>-3</sup>					cm
CTC	avg.	0.23	0.24	0.24	0.29	0.31	30.2
	PBIAS, %	-1.7	-5.9	1.3	-2.5	-1.0	-2.0
	NSE	0.42	0.61	0.67	0.56	0.38	0.65
	RMSD	0.048	0.038	0.030	0.020	0.019	3.0
	RS <sub>p</sub> R	1.66	1.22	0.96	0.75	0.92	0.88
CTE	avg.	0.24	0.25	0.26	0.28	0.30	31.1
	PBIAS, %	-2.2	-6.1	-3.5	1.6	3.4	-1.3
	NSE	0.29	0.47	0.59	0.61	0.39	0.56
	RMSD	0.046	0.041	0.033	0.025	0.026	3.263
	RS <sub>p</sub> R	1.59	1.56	1.18	1.00	1.19	1.03
NTC	avg.	0.24	0.23	0.25	0.29	0.30	30.2
	PBIAS, %	-3.3	-4.6	-4.4	-2.7	0.1	-3.1
	NSE	0.54	0.82	0.65	0.42	-0.04	0.76
	RMSD	0.039	0.024	0.032	0.030	0.020	2.440
	RS <sub>p</sub> R	1.89	0.62	0.93	1.59	1.29	0.81
NTE	avg.	0.25	0.23	0.23	0.28	0.31	31.0
	PBIAS, %	-6.3	-0.4	5.7	3.2	-1.7	-0.3
	NSE	0.51	0.66	0.62	0.51	0.59	0.74
	RMSD	0.043	0.039	0.038	0.030	0.017	2.84
	RS <sub>p</sub> R	2.44	1.64	1.72	0.73	0.56	0.89

† CTC, conventional tillage, conventional management; CTE, conventional tillage, ecological management; NTC, no-till, conventional management; NTE, no-till, ecological management.

‡ PBIAS, percent bias; NSE, Nash–Sutcliffe efficiency; RMSD, root mean squared deviation; RS<sub>p</sub>R, RMSD/pooled sample variance ratio.

## RESULTS AND DISCUSSION

### Model Calibration (2004–2010)

Simulated soil water content and total soil water storage for the calibration plots under the CTC treatment were, in general, acceptable when compared with neutron probe measured values based on the statistics used. Figure 1 illustrates the simulated and measured soil water content and soil water storage from 10 April to 1 November each year. The statistics for soil moisture simulation under all treatments are given in Table 2. Although for some individual soil layers there were a few low NSE values and high RS<sub>p</sub>R values, soil water storage was simulated with PBIAS within  $\pm 5\%$ , NSE > 0.50, and RS<sub>p</sub>R < 1.5. For soil water content in each individual soil layer, the PBIAS was within  $\pm 10\%$  and the NSE was generally close to or greater than 0.50. These statistics were comparable to stand-alone CERES-Wheat under an irrigated condition (Thorp et al., 2010) and were much better than previously reported for CERES-Wheat used to simulate winter wheat under dryland conditions in Colorado (Saseendran et al., 2004).

The simulation of water content in the soil surface layer (0–34 cm) was worse than in deeper soil layers, which might have been due to measurement error of the neutron probe. The neutron probe was placed at a depth of 23 cm for the soil surface layer, but the measured soil water content at this depth cannot represent the moisture in the top 10 cm of soil. Also, many interacting factors affect the simulation of the first soil layer water content, such as surface energy dynamics. In addition to climate effects, the energy dynamics must accurately reflect the effects of residue coverage, which was not measured in this study.

The calibrated crop parameters are given in Table 3. The simulated emergence and maturity dates, although not measured, were close to those found in the literature. The simulated spring wheat growing season from planting to maturity was 107 d when averaged across the 7 yr (2004–2010), close to the observed 106 DAP from field observations on 12 hard red spring wheat cultivars in Montana (Talbert et al., 2001). Our grain-filling duration is not comparable with that listed in Talbert et al. (2001) because in CERES-Wheat the grain fill starts after anthesis, while in Talbert et al. (2001) the grain filling was considered to start at heading; the heading date, however, is not an output of CERES-Wheat. On average, the simulated maturity date was 8 d before the recorded harvest date. The average simulated emergence date was 10 DAP, close to the 11 DAP predicted by NDAWN. The simulated tiller number at harvest was 351 m<sup>-2</sup>, about 5% less than the actual observed harvest tiller number of 368 m<sup>-2</sup> across all the calibration years in 2004 to 2010.

The grain yield simulation was acceptable for the calibration treatment, but the biomass was not predicted satisfactorily based on the calculated statistics. The simulated vs. observed grain yield and total aboveground biomass for the calibration CTC treatment are depicted in Fig. 2. The average simulated yield was 2295 kg ha<sup>-1</sup>, 4.6% lower than the observed yield of 2405 kg ha<sup>-1</sup> from 2004 to 2010. The NSE and RS<sub>p</sub>R values were 0.78 and 0.61, respectively, indicating a good simulation of yield. Although the simulated total aboveground biomass was only 0.7% lower than the observed average biomass of 6378 kg ha<sup>-1</sup> when averaged across the 7 yr, the RS<sub>p</sub>R was <1.5; the NSE, however, was only 0.38, indicating an unacceptable performance in total biomass simulation according to the criteria set in this study. The average simulated harvest index

**Table 3. Calibrated parameter values for spring wheat development, growth, and yield at Sidney, MT, for use with RZWQM2 with CERES-Wheat using the ecotype of DS3585.**

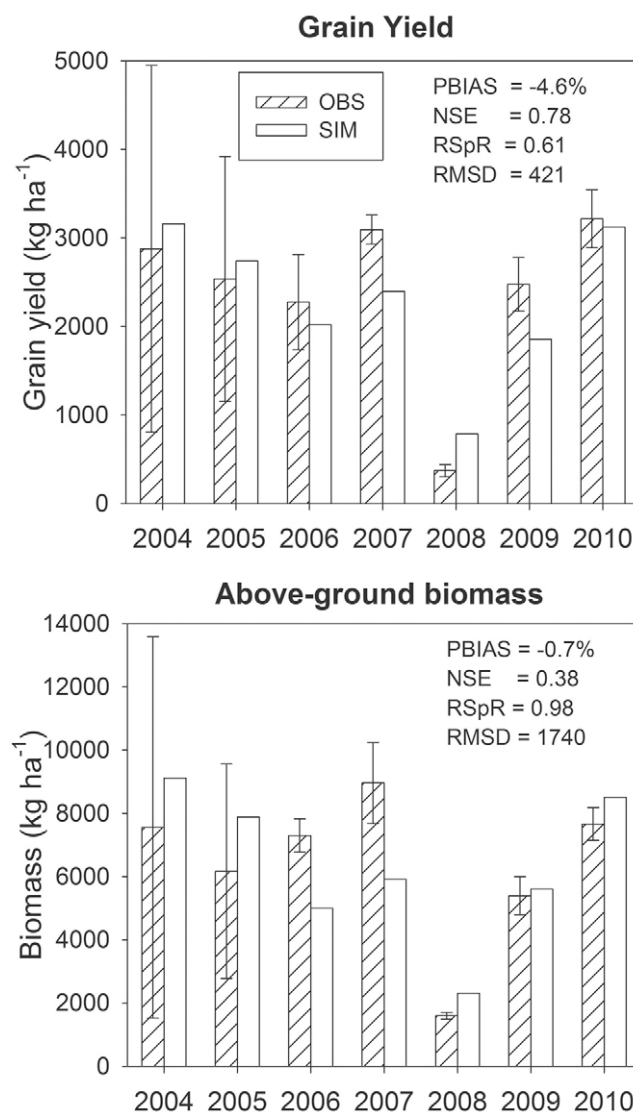
Parameter	Description	Value
<b>Crop development</b>		
P1	duration of phase end juvenile to double ridges, °C d	400
P1V	days at optimum vernalization temperature required to complete vernalization	0
P1D	reduction in development when photoperiod is 10 h less than the threshold (20 h) relative to that at the threshold, %	21
P2	duration of phase double ridges to end leaf growth, °C d	350
P3	duration of phase end leaf growth to end spike growth, °C d	160
P4	duration of phase end spike growth to end grain fill lag, °C d	300
P5	grain-filling phase duration, °C d	400
PHINT	interval between successive leaf tip appearances, °C d	100
<b>Crop growth and yield</b>		
PARUV	photosynthetically active radiation (PAR) conversion to dry matter ratio before last leaf stage, g MJ <sup>-1</sup>	2.2
PARUR	PAR conversion to dry matter ratio after last leaf stage, g MJ <sup>-1</sup>	2.2
G1	kernel number per unit canopy weight at anthesis, no. g <sup>-1</sup>	15.8
G2	standard kernel size under optimum conditions, mg	30
G3	standard nonstressed dry weight (total including grain) of a single tiller at maturity, g	1.5

was 0.36, close to the average observed value of 0.37 across all years. The observed harvest index showed a wider range, however, with the highest value of 0.46 in 2009 and the lowest value of 0.31 in 2006, while the highest simulated harvest index was 0.40 in 2006 and 2007 and the lowest value was 0.33 in 2009.

### Model Evaluation (2004–2010)

The statistics on soil moisture simulation for the evaluation treatments of CTE, NTC, and NTE showed similar trends to those for the calibration treatment, acceptable in total soil water storage although unsatisfactory for some individual soil layers (Table 2). The simulated average growing length of spring wheat for the ecological management treatments was 95 d, 12 d shorter than the conventional management treatment, which was planted 24 d earlier on average. The simulated tiller number was 367 m<sup>-2</sup> at harvest, within 5% error from the observed value of 374 m<sup>-2</sup> for the ecological treatment.

The model performed similarly to the calibration CTC treatment in terms of grain yield and biomass prediction. The yield prediction was acceptable, with PBIAS < 15%, NSE ≥ 0.65, RS<sub>p</sub>R < 1.5, and RMSD < 500 kg ha<sup>-1</sup> (Table 4). In a dryland study in Kansas, Staggenborg and Vanderlip (2005) reported that CERES-Wheat overestimated wheat yield by 10% (RMSD = 1477 kg ha<sup>-1</sup>) and 22% (RMSD = 1439 kg ha<sup>-1</sup>) in wheat–sorghum [*Sorghum bicolor* (L.) Moench]–fallow and wheat–fallow systems, respectively. The CERES-Wheat model did not perform as well under the semiarid dryland conditions in this study as it did in the more humid area of the



**Fig. 2. Observed (OBS) and simulated (SIM) yield and biomass under spring wheat at Sidney, MT, for the calibration conventional tillage and conventional management treatment. Simulations done with RZWQM2 with CERES-Wheat. Error bars represent ±1 standard deviation; PBIAS is percent bias, NSE is Nash–Sutcliffe model efficiency, RSPr is root mean squared deviation to pooled sample variance ratio, and RMSD is root mean squared deviation (kg ha<sup>-1</sup>).**

United Kingdom (Bannayan et al., 2003) or under irrigated conditions in arid Arizona (Thorp et al., 2010). In our study, the performance of the RZWQM2 model in simulating spring wheat yield was comparable to what was found by Bannayan et al. (2003) and Thorp et al. (2010) and better than the models used by Toure et al. (1995), who reported PBIAS ranging from –52 to 43%.

The simulated harvest index was 0.36 averaged across all evaluation treatments, close to the observed values of 0.33, 0.34, and 0.37 for CTE, NTE, and NTC, respectively. In this study, however, the performance of the RZWQM2 model was not acceptable in simulating the total aboveground biomass at harvest, especially in 2006 and 2007. Although the PBIAS values were within ±15%, the RS<sub>p</sub>R values were >1.5 in all evaluation treatments and the NSE values were <0.4, with negative values for CTE and NTE treatments. A consistent

**Table 4. Observed and simulated wheat grain yield for all evaluation treatments (CTE, conventional tillage, ecological management; NTC, no-till, conventional management; NTE, no-till, ecological management) at Sidney, MT. Simulations were done with RZWQM2 with CERES-Wheat.**

Year	CTE		NTC		NTE	
	Observed	Simulated	Observed	Simulated	Observed	Simulated
	kg ha <sup>-1</sup>					
2004	1925 (1366)†	2587	3833 (1003)	3150	3318 (1104)	2573
2005	2310 (264)	2531	2657 (1168)	2731	2687 (375)	2510
2006	1369 (100)	1323	1989 (330)	1974	1562 (131)	1290
2007	1701 (122)	1622	2642 (304)	2329	1840 (300)	1579
2008	167 (25)	708	546 (207)	773	336 (258)	693
2009	2298 (372)	2055	2775 (324)	1797	2549 (254)	1968
2010	1987 (106)	2537	2852 (360)	3100	2131 (191)	2486
Statistics‡						
Avg.	1680 (336)	1909	2471 (528)	2265	2060 (373)	1871
PBIAS, %		13.7		-8.3		-9.2
NSE		0.65		0.73		0.76
RS <sub>p</sub> R		1.20		0.92		1.16
RMSD		405		484		434

† Standard deviations in parentheses.

‡ PBIAS, percent bias; NSE, Nash–Sutcliffe efficiency; RMSD, root mean squared deviation; RS<sub>p</sub>R, RMSD/pooled sample variance ratio.

underestimation of biomass was found in 2006 and 2007 for all treatments, from -18 to -52%. One reason for this underestimation might be the high sampling error due to the limited number (two) of small quadrats in each plot sampling in all years. For example, the total aboveground biomass in the three replicate plots for the CTE were 7547, 2324, and 1328 kg ha<sup>-1</sup> in 2004, leading to a coefficient of variance (ratio of standard deviation to mean) of 89%.

Ma et al. (2011) stated that the performance of a model may be considered acceptable when it responds correctly to the differences among treatments. Although the aboveground biomass simulation was not acceptable according to the statistical analyses used in this study, both simulated yield and biomass responded correctly to the treatment differences. Field data analysis showed that no impact of tillage on yield and biomass was evident, while they were greatly influenced by management, with a 23 and 16% reductions in yield and biomass, respectively, in the ecological management treatments. Average grain yields in the conventional and ecological management systems were 2438 vs. 1870 kg ha<sup>-1</sup> for the field observation and 2280 vs. 1890 kg ha<sup>-1</sup> for the simulation. For grain yield, the simulated reduction due to the ecological treatments was 17.1% compared with the conventional treatments, which was close to the observed reduction of 23.3%. For biomass, the simulated and observed reductions due to ecological treatments were 16.7 and 16.0%, respectively, when compared with the conventional management treatment. The model captured the low yield and biomass in the dry year, which is essential for model application under dryland conditions. Although the low yields were consistently overestimated, the simulated yield and biomass in the dry year of 2008 were much less than those simulated in other years, thus matching the observed trends.

### Water and Temperature Effects (2004–2010)

A useful application of the agricultural system model is to aid in identifying and analyzing cause-and-effect relationships. In this study, plant-available soil water rather

than soil N was found to be a major factor affecting yield. Both field observations and simulations showed high yield in wet years (i.e., 2010) and low yield in dry years (i.e., 2008). The simulation showed no N stress for all years, which was supported by field observations. In 2009 and 2010 when the N application was about 50 kg ha<sup>-1</sup>, approximately 50% of N applied in other years, the observed data indicated that yields at the low N rate in 2009 and 2010 were comparable to the high N rate with a similar rainfall pattern in 2004 and 2005.

The model can also help understand the impact of rainfall and temperature on crop production. Figure 3 depicts the daily rainfall from April to August, cumulative growing degree days since planting, and simulated values of biomass accumulation and phenology in 2004, 2006, and 2008 for the NTC and NTE treatments under no-till as an example. In 2008, the total rainfall during the growing season from April to August was 12.4 cm, about 50% of the long-term average. The extremely low rainfall in 2008 translated to an extremely low biomass accumulation. The rainfall distribution pattern or timing also affected crop growth. During the growing season in 2006, the total rainfall was 24.8 cm, similar to that in 2004. Simulated total aboveground biomass at harvest in 2006 was 54% of that in 2004 for NTC and 46% for NTE, however, due to a long dry period from mid-June to early August in 2006.

Figure 3 also shows temperature effects on the duration of crop growth. The slope of the cumulative growing degree days was steeper in 2006 than in 2004. For example, for the NTC treatment, the cumulative growing degree days on 31 Aug. 2006 was 2434°C d since planting, while in 2004 it was 2024°C d, which was 16.8% lower than in 2006, although the planting date in 2006 was later than in 2004. This indicates a warmer temperature during the growing season in 2006. The average temperature during April through August in 2004 was 14.1°C, while it was 17.5°C in 2006. The 3.4°C increase in temperature in 2006 resulted in a significantly shorter growing period in 2006. For the NTC treatment, the simulated maturity date in 2004 was 114 DAP, while it was 93 DAP in 2006.

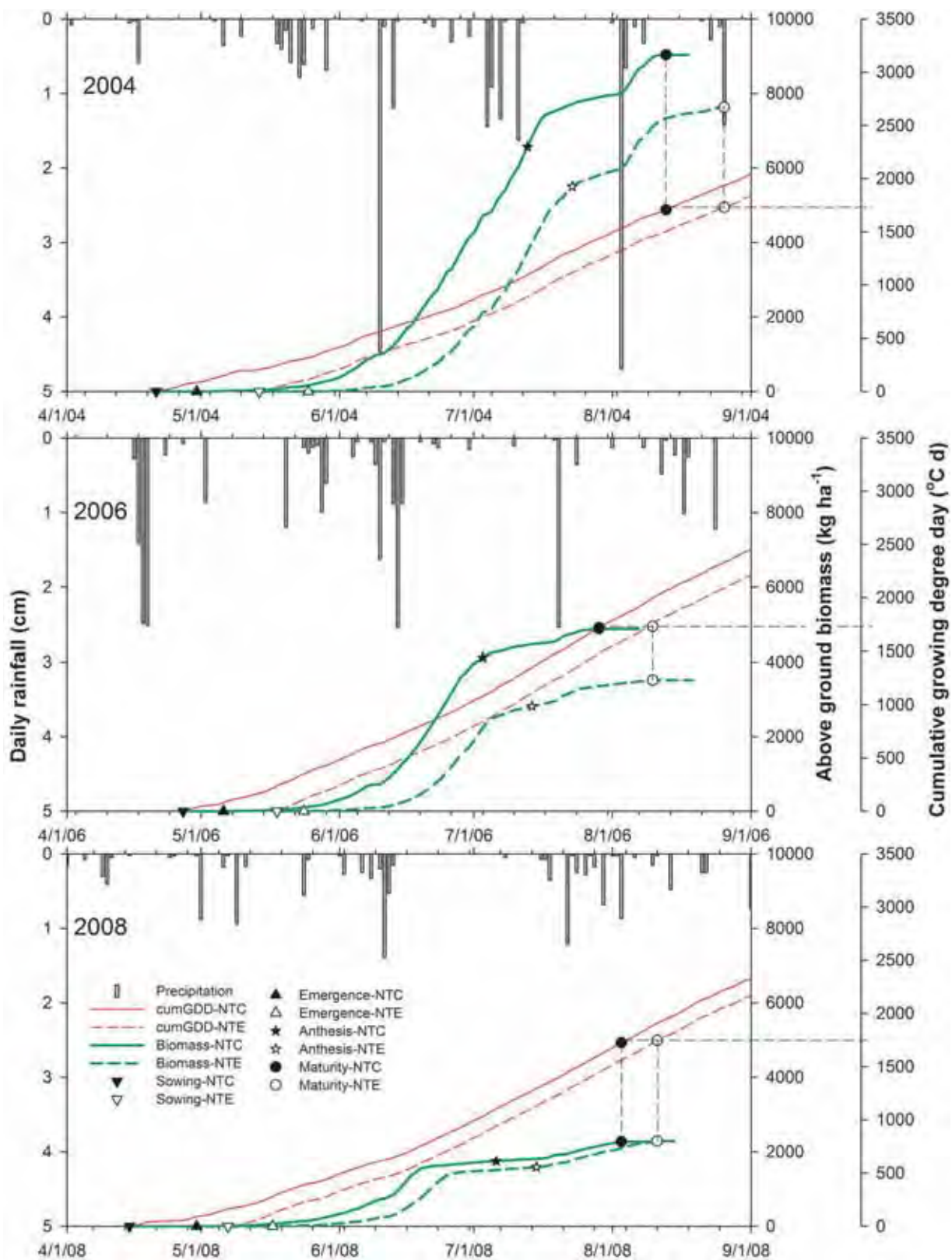


Fig. 3. Daily rainfall during the growing season, growing degree days since planting (cumGDD), simulated biomass accumulation, and essential phenology dates in 2004, 2006, and 2008 for no-till, conventional management (NTC) and no-till, ecological management (NTE) of spring wheat at Sidney, MT. Simulations were done with RZWQM2 with CERES-Wheat.



**Table 5. Simulated long-term (1961–2010) spring wheat grain yield at Sidney, MT. Simulations were done with RZWQM2 with CERES-Wheat for a continuous annual wheat cropping system under conventional tillage, conventional management (CTC), conventional tillage, ecological management (CTE), no-till, conventional management (NTC), and no-till, ecological management (NTE) .**

Years	Yield					Yield difference	
	CTC	CTE	NTC	NTE	Avg.	Ecol. – Conv.	No-till – Tilled
	kg ha <sup>-1</sup>					%	
1961–1970	2125	1934	2089	1895	2011	-9.1	-1.8
1971–1980	2044	1787	2002	1742	1894	-12.8	-2.3
1981–1990	1527	1373	1501	1343	1436	-10.3	-1.9
1991–2000	1803	1997	1775	1950	1881	10.3	-1.9
2001–2010	2006	1707	1993	1661	1842	-15.8	-1.6
Avg.	1901	1760	1872	1718	1813	-7.8	-1.9

**Table 6. Long-term (1961–2010) average of simulated hydrologic components for spring wheat at Sidney, MT. Simulations were done with RZWQM2 with CERES-Wheat (ET is evapotranspiration).**

Treatment†	Precipitation	Runoff	Infiltration	Actual	Actual	Actual	Deep seepage	Potential	Potential	Potential
				evaporation	transpiration	ET		evaporation	transpiration	ET
cm										
January–December										
CTC	34.5	3.0	31.0	8.5	18.7	27.2	4.0	38.0	28.6	66.5
CTE		3.0	31.0	8.9	17.8	26.7	4.5	41.1	28.7	69.8
NTC		3.1	30.9	8.3	18.7	27.0	4.0	34.9	28.6	63.5
NTE		3.2	30.8	8.7	17.8	26.4	4.6	38.2	28.5	66.7
Avg.	34.5	3.1	30.9	8.6	18.3	26.8	4.3	38.1	28.6	66.6
April–August										
CTC	23.5	1.8	22.0	5.2	18.7	23.9	2.9	27.3	28.6	55.8
CTE		1.8	22.0	5.5	17.8	23.3	3.8	29.5	28.7	58.1
NTC		1.9	21.8	5.0	18.7	23.7	2.9	24.9	28.6	53.4
NTE		2.0	21.8	5.4	17.8	23.1	3.9	27.2	28.5	55.7
Avg.	23.5	1.9	21.9	5.3	18.3	23.5	3.4	27.2	28.6	55.8

† CTC, conventional tillage, conventional management; CTE, conventional tillage, ecological management; NTC, no-till, conventional management; NTE, no-till, ecological management.

**Table 7. Simulated yield and hydrologic components in a spring wheat–fallow cropping system compared with a continuous annual spring wheat cropping system during 1961–2010 at Sidney, MT. Simulations were done with RZWQM2 with CERES-wheat.**

Parameter	Wheat–fallow (WF)			Wheat–wheat (WW)	Difference WF – WW		
	Wheat year	Fallow year	Avg.		Absolute	Percentage	
						%	
Yield, kg ha <sup>-1</sup>	2062	0	1031	1813	-782	-43	
Precipitation, cm	34.5	34.5	34.5	34.5	0	0	
Runoff, cm	3.1	3.2	3.2	3.1	0	2	
Infiltration, cm	30.9	30.7	30.8	30.9	0	0	
Actual evaporation, cm	9.6	11.0	10.3	8.6	1.7	20	
Actual transpiration, cm	18.9	0.0	9.5	18.3	-8.8	-48	
Actual evapotranspiration, cm	28.5	11.0	19.8	26.8	-7.0	-26	
Deep seepage, cm	6.9	15.6	11.2	4.3	6.9	161	
Potential evaporation, cm	57.0	61.2	59.1	38.1	21.0	55	
Potential transpiration, cm	29.5	0.0	14.7	28.6	-13.9	-48	
Potential evapotranspiration, cm	86.5	61.2	73.8	66.6	7.2	11	

## Model Application (1961–2010)

### Long-Term Yield

Weather data analysis indicated that the spring growing season in the recent 10 yr was relatively dry and cool at Sidney, MT. April to August precipitation in 2001 to 2010 averaged 21.4 cm, about 2 cm less than the 50-yr average. The average maximum temperature in that decade was 15.7°C, slightly lower than the long-term average (Sainju et al., 2011). In general, compared with long-term averages (1961–2010), the weather was wet and cool in 1961 to 1970, wet and warm in 1991 to 2000, and dry and warm in 1981 to 1990.

The simulated long-term average grain yield of spring wheat was 1813 kg ha<sup>-1</sup> across all treatments from 1961 to 2010

(Table 5). The long-term average yield with conventional management (1887 kg ha<sup>-1</sup>, average of yields in CTC and NTC in Table 5) was higher than the long-term yield previously reported at three sites (1513, 1589, and 1829 kg ha<sup>-1</sup> from 1960–1990) in Saskatchewan, Canada (Chipanshi et al., 1997). The average simulated yield from 1961 to 2010 indicated a minimal impact of tillage, which is similar to findings from the field experiment and short-term simulation for 2004 to 2010. The long-term simulation (1961–2010), however, showed that ecological management led to an average yield reduction of 7.8% compared with conventional management. This percentage was lower than that obtained in the short-term (2004–2010) field experiment (23.3%) and simulation (17.1%).



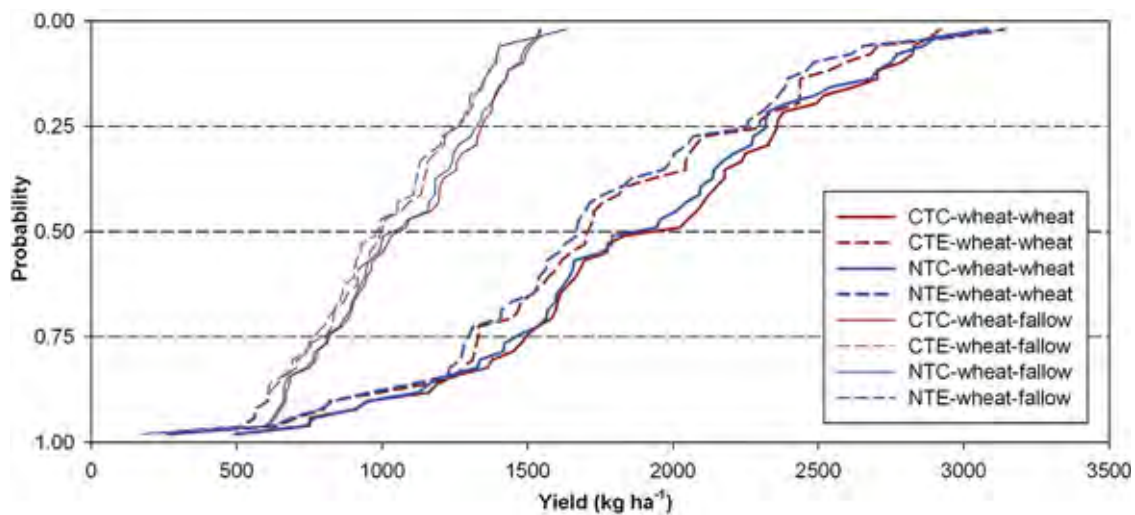


Fig. 4. Probability distribution of simulated long-term (1961–2010) spring wheat yield with wheat–wheat and wheat–fallow cropping systems under different treatments at Sidney, MT. The simulated yield for wheat–fallow was subsequently annualized using the yield in wheat years divided by 2. Simulations done with RZWQM2 with CERES-Wheat. CTC: conventional tillage, conventional management; CTE: conventional tillage, ecological management; NTC: no-till, conventional management; NTE: no-till, ecological management.

The yield difference in management treatments in 2001 to 2010 was close to the short-term simulation, while in other years the yield reduction due to ecological management was less pronounced. A 10.3% yield increase in ecological management treatments during 1991 to 2000 due to high growing-season precipitation was also simulated. The simulated yield was lowest in the driest and warmest years (1981–1990) and highest in the wettest and coolest years (1961–1970).

For this continuous spring wheat system, the probability of obtaining a given simulated low yield (<1200 kg ha<sup>-1</sup>) was not greatly different among all treatments. The probability of obtaining a given simulated high yield (1200–2800 kg ha<sup>-1</sup>), however, was consistently higher for conventional management than for ecological management (Fig. 4). This suggests that during very dry years, a management effect on wheat yield is not evident; for the average and wet years, the simulated yield under the conventional treatment significantly exceeded that under the ecological treatment.

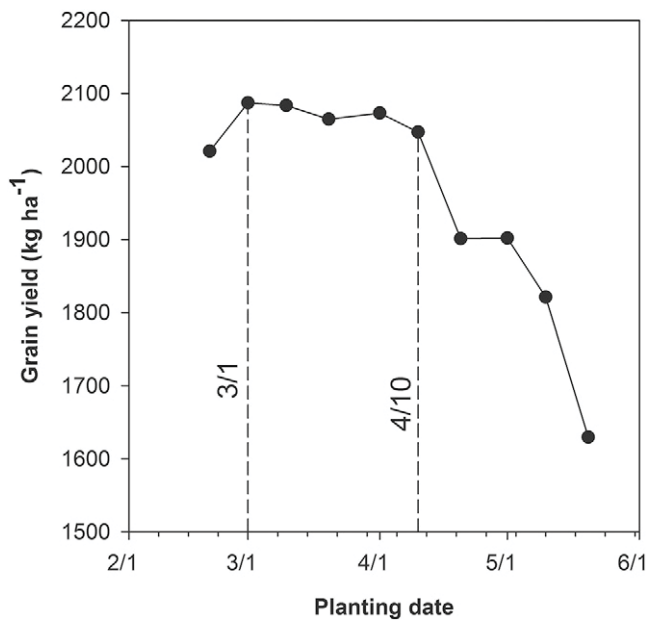
The analysis showed a large deficit in the water supply to spring wheat under this dryland condition (Table 6). The calculated potential evapotranspiration (ET) during the spring growing season (April–August) was 55.8 cm, while the simulated water consumption of spring wheat during this period was 23.5 cm, accounting for 42% of the potential water need. This suggests that, to meet 100% of the potential ET, the spring wheat should be irrigated with an additional 32.3 cm of water during the growing season in this region. The model-simulated total water consumption by wheat and water loss through seepage and runoff exceeded the total precipitation by 5.3 cm, indicating a soil water storage decrease during the growing season. In other words, 5.3 cm of water used by the crop ET was supplied by soil water storage, which accounted for 23% of the water use during the growing season. The decreasing trend of soil water storage during the spring wheat growing season was seen consistently in the recent field experiment years (Fig. 1). The annual water balance in Table 6 indicates that the actual ET (23.5 cm) of spring wheat during

the growing season was about 70% of the annual precipitation, while the remaining 30% of precipitation was lost through soil evaporation, deep seepage, and surface runoff.

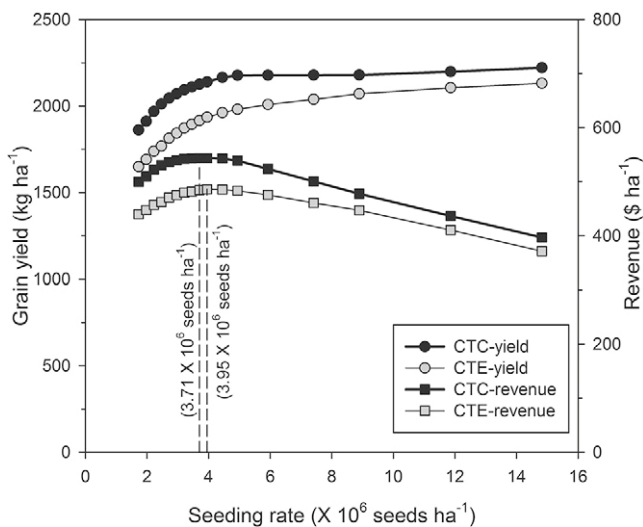
#### Wheat–Fallow vs. Wheat–Wheat

The long-term simulation indicated that the wheat–fallow system would increase wheat grain yield in the wheat growing years. The long-term average yield across all treatments when fallowed every other year was 2062 kg ha<sup>-1</sup> (Table 7), 249 kg ha<sup>-1</sup> (13.7%) higher than the long-term average yield of 1813 kg ha<sup>-1</sup> (Table 5) in the continuous wheat system. Because there was no crop grown in the fallow years, however, the annualized total average yield was half of the yield in wheat years. This translates to an annualized total average yield of 1031 kg ha<sup>-1</sup> for the wheat–fallow system, 43.1% (782 kg ha<sup>-1</sup>) less than the annual yield for the continuous wheat–wheat system. This is close to the results of a simulation study for Swift Current, SK, Canada, where the simulated 25-yr cumulative spring wheat yield for wheat–fallow was 45.0% less than continuous wheat (Kersebaum et al., 2008; 34.5 vs. 23.8 Mg ha<sup>-1</sup>). The probability analysis for the simulated long-term yield of wheat–fallow under different treatments showed a similar trend to the wheat–wheat system (Fig. 4). The probability of obtaining a given yield was much lower for the wheat–fallow than the wheat–wheat system due to fallow every other year.

The comparison of simulated values for the hydrology components suggests that fallow every other year contributed little to plant-available water in the following year under the dryland conditions in Sidney, MT. Of the 34.5 cm of precipitation in the fallow year, 9.4% (3.2 cm) was lost to surface runoff, 31.9% (11.0 cm) to soil evaporation, and 45.2% (15.6 cm) to deep percolation (Table 7). Only 12.1% (4.2 cm) of the total precipitation was carried over to the next wheat year. The simulated actual crop ET was 28.5 cm in the wheat growing year after 1 yr of fallow, 1.7 cm higher than the 26.8 cm simulated for the continuous wheat rotation



**Fig. 5.** Simulated average spring wheat grain yield during 1961 to 2010 at various planting dates at Sidney, MT. Simulations were done with RZWQM2 with CERES-Wheat.



**Fig. 6.** Average spring wheat grain yield and economic return (revenue = crop income minus seed cost) at various seeding rates in 1961 to 2010 for the conventional tillage, conventional management (CTC) and conventional tillage, ecological management (CTE) treatments as simulated by RZWQM2 with CERES-Wheat, assuming a seed weight of 30 mg kernel<sup>-1</sup>, a seed price of US\$0.51 kg<sup>-1</sup> (US\$14.2 bu<sup>-1</sup>), and a wheat grain selling price of US\$0.28 kg<sup>-1</sup> (US\$7.7 bu<sup>-1</sup>) based on the local prices reported for March 2011 by the National Agricultural Statistics Service (2011b).

(Table 6). Therefore, of the 4.2 cm of water carried over from the fallow year, only 1.7 cm was actually used by the crop. This analysis indicates that a 1-cm increase in water consumption by spring wheat increased the yield by 146 kg ha<sup>-1</sup> (2062 vs. 1813 kg ha<sup>-1</sup>). This relationship is close to the findings in field experiments conducted by Brown et al. (1981; 135 kg ha<sup>-1</sup> cm<sup>-1</sup>) and Nielsen et al. (2002; 140 kg ha<sup>-1</sup> cm<sup>-1</sup>) in the northern and central Great Plains. The rest of the

carryover water (2.5 cm) from the fallow years was lost through deep seepage in the wheat growth year. This, in combination with lower annualized yields, makes the wheat–fallow system less efficient from a productivity standpoint.

### Optimum Planting Date and Seeding Rate

The simulation of grain yield with long-term weather data (1961–2010) at various planting dates suggests that the optimum planting dates should be between 1 March and 10 April (Fig. 5). The average simulated yield within this planting window was high and stable, from 2047 to 2087 kg ha<sup>-1</sup>. If wheat is planted after 20 April, the grain yield drops dramatically, from 2047 kg ha<sup>-1</sup> for planting on 10 April to 1629 kg ha<sup>-1</sup> for planting on 20 May. It should be noted that although the simulated spring wheat yield planted between 1 March and 20 April showed a higher yield than later planting dates, field trafficability should be taken into account because soil moisture was higher during the early spring and soils typically would be frozen. The traditional planting date for spring wheat in this region has been mid- to late April, which is outside the optimum planting date simulated by the RZWQM2 model. This simulation study shows that farmers should plant wheat 2 to 4 wk earlier, if possible, to secure an optimum yield.

The simulation also indicated that crop revenue reached a maximum at a seeding rate of  $3.71 \times 10^6$  seeds ha<sup>-1</sup> for conventional management and  $3.95 \times 10^6$  seeds ha<sup>-1</sup> for ecological management. Although the crop yield kept increasing with increasing seeding rate, the cost of seed increased as well, which reduced marginal revenue (Fig. 6). Of note is that in Fig. 6 the calculation of revenue was the difference between crop yield income and seed cost. Other capital costs such as machinery and labor were not considered because we assumed them to be fixed under various seeding rates. This suggests that the current seeding rate used for the conventional management ( $2.22 \times 10^6$  seed ha<sup>-1</sup>) could be increased to target higher revenue. This calculation was based on prices for March 2011 (National Agricultural Statistics Service, 2011b; US\$0.28 kg<sup>-1</sup> for wheat grain and US\$ 0.52 kg<sup>-1</sup> for wheat seed); however, the optimum seeding rate may change according to fluctuations in market input costs and output returns.

## CONCLUSIONS

The RZWQM2 model simulated soil water and crop yield to an acceptable level for the dryland spring wheat under various tillage methods, planting dates, and seeding rates in 2004 to 2010 at the Rasmussen site near Sidney, MT. Although the experimental observed total aboveground biomass from sampling of two small quadrats was poorly predicted, the model captured differences between conventional and ecological treatments. The model showed no impact of tillage but indicated that late planting greatly reduced grain yield and biomass under ecological management, in agreement with the observed differences among treatments. The simulation also showed that, aside from growing season rainfall, the rainfall distribution and small differences in air temperature may significantly affect crop yield. This finding suggests that global changes in rainfall and temperature may lead to significant

yield losses in the future. The simulated long-term yield loss in the ecological treatments compared with conventional management, which may be mainly due to late planting, was much less than the short-term observed and simulated yield losses. It confirms the significance of using models and long-term weather data to extend the results of limited years of field experimentation. Such an extension requires that the model be calibrated and validated on the original field experiment. The spring wheat–fallow system was simulated to be less productive than continuous spring wheat because the contribution of rainfall in the fallow year to the following wheat year was very small. Continuous cropping systems, however, are prone to weed infestation and require mechanical tillage or chemical weed control to maintain productivity. Finally, the model also provided long-term optimal planting dates and seeding rates for this area.

### ACKNOWLEDGMENTS

We would like to express our appreciation to Dr. Gregory McMaster at the USDA-ARS Agricultural Systems Research Unit in Fort Collins, CO, and Dr. Jeffery White at the USDA-ARS Arid Land Agricultural Research Center in Maricopa, AZ, for their help in understanding wheat physiology and the CERES-Wheat model. We are thankful to Dr. David Archer at the USDA-ARS Northern Great Plains Research Lab. for providing information for the economic analysis. Dr. Jeffery White, Dr. David Archer, and Dr. David Nielsen are acknowledged for performing the ARS internal review.

### REFERENCES

- Ahuja, L.R., K.W. Rojas, J.D. Hanson, M.J. Shaffer, and L. Ma, editors. 2000. Root Zone Water Quality Model: Modeling management effects on water quality and crop production. Water Resour. Publ., Highlands Ranch, CO.
- Allen, R.G., L.S. Pereira, D. Raes, and M. Smith. 1998. Crop evapotranspiration: Guidelines for computing crop water requirements. Irrig. Drain. Pap. 56. FAO, Rome.
- Bannayan, M., N.M.J. Crout, and G. Hoogenboom. 2003. Application of the CERES-Wheat model for within-season prediction of winter wheat yield in the United Kingdom. *Agron. J.* 95:114–125. doi:10.2134/agronj2003.0114
- Bouyoucos, G.J. 1962. Hydrometer method improved for making particle size analyses for soils. *Agron. J.* 54:464–465. doi:10.2134/agronj1962.00021962005400050028x
- Brooks, R.J., and A.T. Corey. 1964. Hydraulic properties of porous media. *Hydrol. Pap. 3*. Colorado State Univ., Fort Collins.
- Brown, P.L., A.L. Black, C.M. Smith, J.W. Enz, and J.M. Caprio. 1981. Soil water guidelines and precipitation probabilities for barley and spring wheat in flexible cropping systems in Montana and North Dakota. *Montana Coop. Ext. Bull.* 356. Montana State Univ., Bozeman.
- Chipanshi, A.C., E.A. Ripley, and R.G. Lawford. 1997. Early prediction of spring wheat yields in Saskatchewan from current and historical weather data using the CERES-Wheat model. *Agric. For. Meteorol.* 84:223–232. doi:10.1016/S0168-1923(96)02363-5
- Chanasyk, D.S., and M.A. Naeth. 1996. Field measurement of soil moisture using neutron probes. *Can. J. Soil Sci.* 76:317–323. doi:10.4141/cjss96-038
- Deibert, E.J., E. French, and B. Hoag. 1986. Water storage and use by spring wheat under conventional tillage and no-till in continuous and alternate crop–fallow systems in the northern Great Plains. *J. Soil Water Conserv.* 41(1):53–58.
- Fenster, C.R. 1973. Stubble mulching. In: *Conservation tillage: A handbook for farmers*. Soil Conserv. Soc. Am., Ankeny, IA. p. 29–34.
- Godwin, D., J. Ritchie, U. Singh, and L. Hunt. 1990. A user's guide to CERES Wheat, V2.10. 2nd ed. Int. Fert. Dev. Ctr., Muscle Shoals, AL.
- Hanson, J.D., K.W. Rojas, and M.J. Shaffer. 1999. Calibrating the Root Zone Water Quality Model. *Agron. J.* 91:171–177. doi:10.2134/agronj1999.00021962009100020002x
- Jones, J.W., G. Hoogenboom, C.H. Porter, K.J. Boote, W.D. Batchelor, L.A. Hunt, et al. 2003. The DSSAT cropping system model. *Eur. J. Agron.* 18:235–265. doi:10.1016/S1161-0301(02)00107-7
- Kersebaum, K.C., A. Wurbs, R. de Jong, C.A. Campbell, J. Yang, and R.P. Zentner. 2008. Long-term simulation of soil–crop interactions in semiarid southwestern Saskatchewan, Canada. *Eur. J. Agron.* 29:1–12. doi:10.1016/j.eja.2008.01.011
- Lenssen, A.W., G.D. Johnson, and G.R. Carlson. 2007. Cropping sequence and tillage system influences annual crop production and water use in semiarid Montana, USA. *Field Crops Res.* 100:32–43. doi:10.1016/j.fcr.2006.05.004
- Ma, L., L.R. Ahuja, B.T. Nolan, R.W. Malone, T.J. Trout, and Z. Qi. 2012. Root Zone Water Quality Model (RZWQM2): Model use, calibration, and validation. *Trans. ASABE* 55:1425–1446.
- Ma, L., L.R. Ahuja, S.A. Saseendran, R.W. Malone, T.R. Green, B.T. Nolan, et al. 2011. A protocol for parameterization and calibration of RZWQM2 in field research. In: L.R. Ahuja and L. Ma, editors, *Methods of introducing system models into agricultural research*. *Adv. Agric. Syst. Model.* 2. ASA, CSSA, and SSSA, Madison, WI. p. 1–64.
- Ma, L., G. Hoogenboom, L.R. Ahuja, J.C. Ascough II, and S.A. Saseendran. 2006. Evaluation of the RZWQM–CERES–Maize hybrid model for maize production. *Agric. Syst.* 87:274–295. doi:10.1016/j.agry.2005.02.001
- Ma, L., G. Hoogenboom, L.R. Ahuja, D.C. Nielsen, and J.C. Ascough II. 2005. Development and evaluation of the RZWQM–CROPGRO hybrid model for soybean production. *Agron. J.* 97:1172–1182. doi:10.2134/agronj2003.0314
- Ma, L., R.W. Malone, P. Heilman, D.B. Jaynes, L.R. Ahuja, S.A. Saseendran, et al. 2007. RZWQM simulated effects of crop rotation, tillage, and controlled drainage on crop yield and nitrate–N loss in drain flow. *Geoderma* 140:260–271. doi:10.1016/j.geoderma.2007.04.010
- Malone, R.W., L. Ma, P. Heilman, D.L. Karlen, R.S. Kanwar, and J.L. Hatfield. 2007. Simulated N management effects on corn yield and tile-drainage nitrate loss. *Geoderma* 140:272–283. doi:10.1016/j.geoderma.2007.04.011
- McMaster, G.S., J.F. White, L.A. Hunt, P.D. Jamieson, S.S. Dhillon, and J.I. Ortiz-Monasterio. 2008. Simulating the influence of vernalization, photoperiod, and optimum temperature on wheat developmental rates. *Ann. Bot.* 102:561–569. doi:10.1093/aob/mcn115
- Moriasi, D.N., J.G. Arnold, M.W. Van Liew, R.L. Binger, R.D. Harmel, and T.L. Veith. 2007. Model evaluation guidelines for systematic qualification of accuracy in watershed simulations. *Trans. ASABE* 50:885–900.
- Moulin, A.P., and H.J. Beckie. 1993. Evaluation of the CERES and EPIC models for predicting spring wheat grain yield over time. *Can. J. Plant Sci.* 73:713–719. doi:10.4141/cjps93-093
- Nash, J.E., and J.V. Sutcliffe. 1970. River flow forecasting through conceptual models: I. A discussion of principles. *J. Hydrol.* 10:282–290. doi:10.1016/0022-1694(70)90255-6
- National Agricultural Statistics Service. 2007. The census of agriculture. NASS, Washington, DC. [http://www.agcensus.usda.gov/Publications/2007/Full\\_Report/](http://www.agcensus.usda.gov/Publications/2007/Full_Report/) (accessed 27 Jan. 2012).
- National Agricultural Statistics Service. 2011a. Acreage. Released 30 June 2011. NASS, Washington, DC. <http://usda.mannlib.cornell.edu/usda/nass/Acre/2010s/2011/Acre-06-30-2011.pdf>.
- National Agricultural Statistics Service. 2011b. Agricultural prices. Released 29 Apr. 2011. NASS, Washington, DC. <http://usda01.library.cornell.edu/usda/nass/AgriPric/2010s/2011/AgriPric-04-29-2011.pdf>.
- Nielsen, D.C., and M.F. Vigil. 2010. Precipitation storage efficiency during fallow in wheat–fallow systems. *Agron. J.* 102:537–543. doi:10.2134/agronj2009.0348
- Nielsen, D.C., M.F. Vigil, R.L. Anderson, R.L. Bowman, J.G. Benjamin, and A.D. Halvorson. 2002. Cropping system influence on planting water content and yield of winter wheat. *Agron. J.* 94:962–967. doi:10.2134/agronj2002.0962
- Nielsen, D.C., M.F. Vigil, and J.G. Benjamin. 2011. Evaluating decision rules for dryland rotation crop selection. *Field Crops Res.* 120:254–261. doi:10.1016/j.fcr.2010.10.011

- Qi, Z., M.J. Helmers, R.W. Malone, and K.R. Thorp. 2011. Simulating long-term impacts of winter rye cover crop on hydrologic cycling and nitrogen dynamics for a corn–soybean crop system. *Trans. ASABE* 54:1575–1588.
- Ritchie, J.T., U. Singh, D.C. Godwin, and W.T. Bowen. 1998. Cereal growth, development and yield. In: G.Y. Tsuji et al., editors, *Understanding options for agricultural production*. Kluwer Acad. Publ., Dordrecht, the Netherlands. p. 79–98.
- Sainju, U.M., A.W. Lenssen, T. Caesar-TonThat, J.D. Jabro, R.T. Lartey, R.G. Evans, and B.L. Allen. 2011. Dryland residue and soil organic matter as influenced by tillage, crop rotation, and cultural practices. *Plant Soil* 338:27–41. doi:10.1007/s11104-010-0403-5
- Saseendran, S.A., L. Ma, D.C. Nielsen, M.F. Vigil, and L.R. Ahuja. 2005. Simulating planting date effects on corn production using RZWQM and CERES-Maize models. *Agron. J.* 97:58–71. doi:10.2134/agronj2005.0058
- Saseendran, S.A., D.C. Nielsen, L. Ma, L.R. Ahuja, and A.D. Halvorson. 2004. Modeling nitrogen management effects on winter wheat production using RZWQM and CERES-Wheat. *Agron. J.* 96:615–630. doi:10.2134/agronj2004.0615
- Saseendran, S.A., D.C. Nielsen, L. Ma, L.R. Ahuja, and M.F. Vigil. 2010. Simulating alternative dryland rotational cropping systems in the central Great Plains with RZWQM2. *Agron. J.* 102:1521–1534. doi:10.2134/agronj2010.0141
- Schaap, M.G., F.J. Leij, and M.Th. van Genuchten. 1998. Neural network analysis for hierarchical prediction of soil water retention and saturated hydraulic conductivity. *Soil Sci. Soc. Am. J.* 62:847–855. doi:10.2136/sssaj1998.03615995006200040001x
- Schillinger, W.F., S.E. Schofstoll, and J.R. Alldredge. 2008. Available water and wheat grain yield relations in a Mediterranean climate. *Field Crops Res.* 109:45–49. doi:10.1016/j.fcr.2008.06.008
- Shaffer, M.J., K.W. Rojas, D.G. Decoursey, and C.S. Hebson. 2000. Nutrient chemistry processes—OMNI. In: L.R. Ahuja et al., editors, *Root Zone Water Quality Model: Modeling management effects on water quality and crop production*. Water Resour. Publ., Highlands Ranch, CO. p. 119–144.
- Sherman, J.D., D.K. Weaver, M.L. Hofland, S.E. Sing, M. Buteler, S.P. Lanning, et al. 2010. Identification of novel QTL for sawfly resistance in wheat. *Crop Sci.* 50:73–86. doi:10.2135/cropsci2009.03.0145
- Staggenborg, S.A., and R.L. Vanderlip. 2005. Crop simulation models can be used as dryland cropping systems research tools. *Agron. J.* 97:378–384. doi:10.2134/agronj2005.0378
- Talbert, L.E., S.P. Lanning, R.L. Murphy, and J.M. Martin. 2001. Grain fill duration in twelve hard red spring wheat crosses: Genetic variation and association with other agronomic traits. *Crop Sci.* 41:1390–1395. doi:10.2135/cropsci2001.4151390x
- Tanaka, D.L., and J.K. Aase. 1987. Fallow method influences on soil water and precipitation storage efficiency. *Soil Tillage Res.* 9:307–316. doi:10.1016/0167-1987(87)90056-0
- Thorp, K.R., D.J. Hunsaker, A.N. French, J.W. White, T.R. Clarke, and P.J. Pinter, Jr. 2010. Evaluation of the CSM-CROPSIM-CERES-Wheat model as a tool for crop water management. *Trans. ASABE* 53:87–102.
- Thorp, K.R., R.W. Malone, and D.B. Jaynes. 2007. Simulating long-term effects of nitrogen fertilizer application rates on corn yield and nitrogen dynamics. *Trans. ASABE* 50:1287–1303.
- Toure, A., D.J. Major, and C.W. Lindwall. 1995. Comparison of five wheat simulation models in southern Alberta. *Can. J. Plant Sci.* 75:61–68. doi:10.4141/cjps95-010



# MODELING OF SUBSURFACE TILE DRAINAGE USING MIKE SHE

X. Zhou, M. Helmers, Z. Qi

**ABSTRACT.** *Accurate estimation of subsurface drainage flow is essential for effectively evaluating the performance of management strategies in tile-drained landscapes. The objectives of this study were to calibrate and validate MIKE SHE for simulating subsurface tile drainage flow in central Iowa and to evaluate the simulated impact of two specific management strategies on tile flow. The model was calibrated and validated using the measured daily drainage in a 15.2 × 38 m row-cropped plot during 2006-2009 with a split-sample method. Drainage time constant and macropore flow were found to be important in predicting drainage flow. The calibrated drainage time constant of 5.6 h was outside the recommended range, which might be attributed to the shorter response time of subsurface flow at the plot scale used in this study. The calibrated model showed a satisfactory performance in simulating daily tile drainage flow with Nash-Sutcliffe model coefficient values of 0.78 and 0.73 for the calibration and validation periods, respectively. The measured and predicted total drainage from 2006-2009 were 865 and 958 mm, respectively. The results suggest that MIKE SHE has potential for predicting tile drainage flows; guiding management decisions and for assessment of drainage design of water flow to downstream water bodies.*

**Keywords.** *Water management, Drainage time constant, Macropore flow.*

**A**rtificial subsurface drainage systems, primarily tile drains, have been widely used in Iowa and the Midwest United States to remove excess water from agricultural fields for improving the productivity of poorly drained soils. In Iowa, approximately 3.6 million ha of cropland are artificially drained, accounting for more than 25% of the state's agricultural land (Baker et al., 2004). Subsurface drainage allows for greater soil aeration, possibly earlier planting, and better field conditions for crop growth and field operations.

The overall environmental impact of artificial subsurface drainage systems continues to generate significant discussion. Subsurface drainage has the potential to reduce surface runoff and associated loss of surface runoff pollutants including sediment and chemicals (Bengtson et al., 1995; Skaggs et al., 1995). In comparison with undrained soils in the United States, Skaggs et al. (1994) estimated that subsurface drainage reduced surface runoff by 34% to 55%, by altering hydrology through changes in infiltration rate (Shipitalo et al., 2004), water storage capacity (Irwin and Whitely, 1983), soil structure (Gardner et al., 1994), and subsurface water pathways (Gardenas

et al., 2006). Tile drained soils can also act as a buffer for rainfall and thus significantly reduce peak flow volumes (Skaggs and Broadhead, 1982). On the other hand, during the last decade there has been increased concern on the role of artificial subsurface drainage in flood generation and pollutant transport. There has been an increased frequency of severe river floods in the Midwest and other states in the United States during the past decades (Karl et al., 2009). A number of factors, such as climate change, aggressive tillage operations, and changes in land use and cropping systems, could be responsible for the changes in flood behavior. However, it is widely recognized that a better understanding of the role of subsurface drainage in flood behavior is critical (Wiskow and van der Ploeg, 2003), particularly in the tile-drained Corn Belt area. Furthermore, subsurface drainage provides a major transport pathway for soluble agricultural chemicals such as nitrate, which has been identified as a major contributor to the hypoxic zone in the Gulf of Mexico (Rabalais et al., 2001). Generally, nitrate loss increases as subsurface drainage flow volume increases (Cambardella et al., 1999; Kanwar et al., 2005). Therefore, characterizing the response of subsurface drainage flow to precipitation is essential for understanding the hydrologic footprint of land management practices in the Midwest United States.

Field monitoring of subsurface drainage flow is generally expensive and hence many subsurface drainage systems are unengaged. Alternatively, hydrologic models have been developed to simulate underlying hydrological processes and provide a quantitative estimation of drainage discharge. Examples of these models are DRAINMOD, ADAPT, and MIKE SHE. The European Hydrological System model or MIKE SHE was developed by three

---

Submitted for review in December 2011 as manuscript number SW 9568; approved for publication by the Soil & Water Division of ASABE in July 2013.

The authors are **Xiaobo Zhou**, Assistant Scientist, **Matthew Helmers**, **ASABE Member**, Associate Professor, Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, Iowa; and **Zhiming Qi**, **ASABE Member**, Assistant Professor, McGill University, Department of Bioresource Engineering, Sainte-Anne-de-Bellevue, Quebec, Canada. **Corresponding author:** Xiaobo Zhou, Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, IA 50011; phone: 515-294-1230; email: xzhou@iastate.edu.



European Organizations (Danish Hydraulic Institute, British Institute of Hydrology, and the French consulting company SOGREAH) and has been successfully applied in various hydrologic studies, including irrigation planning and management of water resources (Singh et al., 1997), groundwater management (Demetriou and Punthakey, 1999), floodplain studies (Jaber and Shukla, 2005), and land planning and management (Thompson et al., 2004; Helmers et al., 2005).

Application of MIKE SHE to subsurface drainage is limited compared to its application in surface hydrology. The impact of subsurface drainage on pumped discharge and soil moisture storage was simulated at a small agricultural catchment in the United Kingdom using the MIKE SHE model (Al-Khudhairy et al., 1999). They found that theoretically removing subsurface drainage increased the magnitude of flood peaks during the spring recession but reduced annual discharges. These findings are consistent with DRAINMOD studies which have generally found that increased subsurface drainage intensity increases annual discharge from drained catchments (Konyha et al., 1992). Feyen et al. (2000) calibrated and validated MIKE SHE using daily discharge of a 600 km<sup>2</sup> catchment in Belgium. They found that the drainage time constant, a key parameter for subsurface drainage, influenced the peaks of the hydrograph, while the drainage depth had a less pronounced effect on streamflow discharge. This was in agreement with the findings of another study in a tropical mountainous watershed (Sahoo et al., 2006). In all of those studies, however, the MIKE SHE model was calibrated and validated against stream discharge at watershed outlets, and the component of subsurface drainage was incorporated in the model mainly for the purpose of better correlation between the observed and predicted discharges. Consequently, subsurface drainage discharge was not directly measured and could not be compared to model predictions during calibration and validation. Thus, little is known about the accuracy of subsurface drainage predictions of MIKE SHE. Iowa soils are extensively tile drained; there is a clear need to assess the performance of MIKE SHE for simulating subsurface drainage flow. The objectives of this study are: (1) to calibrate and validate MIKE SHE for predicting subsurface drainage flow within a row-cropped setting in central Iowa, and (2) to evaluate the impact of management strategies, including converting row-crops to pasture or adopting shallow drainage, on subsurface drainage flow using MIKE SHE.

## METHODS

### SITE DESCRIPTION

The study location is at the Agricultural Drainage Water Quality - Research and Demonstration Site (ADWQ-RDS) in Pocahontas County, north central Iowa. It is within the Des Moines Lobe region (MLRA 103), which is a nearly level to gently rolling till plain. Subsurface drainage tiles have been widely installed to lower the water table during the spring in this region. Corn and soybeans are the major crops, with some cropland used for hay. The glacial till

soils of the ADWQ-RDS are naturally poorly to somewhat poorly drained.

This research site was established in the late 1980s for a long-term drainage and nutrient management study. The total research area is 4.5 ha with slopes varying between 0.5% and 1.5%. There are a total of 73 individually drained plots at the site. The size of each plot is 38 m in length and 15.2 m in width. At each plot, corrugated plastic drain tiles were installed in 1988 at the center and plot borders parallel to the long dimension (7.6 m spacing) at a depth of 1.06 m. Lateral subsurface flow from adjacent plots was minimized by installing two border drains, which have an outlet to the surface at a remote location. Drainage flow was monitored in the center drainage line of each plot using an automatic pumping and volume monitoring system. The tile flow was generally monitored from April to November to avoid damage during freezing conditions in the winter. More details of the tile drainage design can be found in Lawlor et al. (2008).

Plot No. 7-1 was selected for modeling subsurface tile drainage to minimize the lateral effect from neighboring plots, due to its relatively higher elevation and being separated from the upper land by an open ditch. The predominant soil in this plot is Canisteo clay loam (fine-loamy, mixed, superactive, calcareous, mesic Typic Endoaquolls). The slope of the plot is about 1% in the north-south direction (parallel to tile lines) and nearly flat in the east-west direction. A 2-m digital elevation model (DEM) was created from the LiDAR data of the study area using ArcGIS (ESRI, Redlands, Calif.) to represent the topography of the plot.

### THE MIKE SHE MODEL

The MIKE SHE model is a deterministic, fully distributed and physically-based model that allows for simulation of the major processes occurring in the land phase of the hydrologic cycle (Refsgaard and Storm, 1995; DHI, 2004). It is a computationally intensive model simulating surface, subsurface, and stream flow separately for distributed grid-points using numerical solutions of partial differential equations. The model's distributed nature allows for spatially varying climate variables, vegetation, soil properties, and land uses. MIKE SHE has a module structure consisting of several modules such as the Water Movement (WM) module, Advection/Dispersion of Solutes (AD) module, and Soil Erosion (SE) module, and other modules.

The WM module used in this study consists of several components to represent the major physical processes of the hydrological cycle. These include interception/evapotranspiration, overland and channel flow, unsaturated zone, saturated zone, and snowmelt. The interception/evapotranspiration processes are calculated using meteorological and vegetative data based on the Kristensen-Jensen method (1975). The potential evaporation rate is required as input data to calculate actual evapotranspiration. The 1-D channel flow and 2-D overland flow are modeled using a finite difference approximation of the Saint Venant equations. MIKE SHE uses the 1-D Richards' equation for unsaturated zone flow and the 3-D

Boussinesq equation for saturated zone flow. In MIKE SHE, snowmelt is calculated using the empirical degree-day approach.

## MODEL CONSTRUCTION

### Meteorological Data

Rainfall intensity and reference evapotranspiration ( $ET_0$ ) were required by MIKE SHE as climate input data. The 5-min meteorological data including rainfall, air temperature, solar radiation, relative humidity, and wind speed were collected from an automatic meteorological weather station at the study site. An additional tipping-bucket rain gauge (Campbell Scientific, Inc., Logan, Utah) was installed at the site as a backup system. The FAO Penman-Monteith method was used to estimate  $ET_0$  (FAO, 1998). Snow data was also obtained from the National Climate Data Center (NCDC) station. In MIKE SHE, the degree-day coefficient and threshold melting temperature were set to 2 mm/day/°C and 0°C, respectively. The 30-year (1971-2000) normal precipitation for Pocahontas County was calculated based on the records from the NCDC station at Humboldt, Iowa, located 15 km to the east of the site.

### Land Use

A corn-soybean rotation system was implemented at the modeling plot, with corn planted in odd years (2005, 2007, and 2009) and soybean in even years (2006 and 2008). Since 2004, rye (*Secale cereale*) was planted each year as a winter cover crop to reduce soil and nutrient loss, and was killed in the spring of the following year. Glyphosate resistant corn (*Zea maize*) and soybean (*Glycine max*) were planted after the growth of rye was terminated by glyphosate in early spring. The plot was fall disked and field cultivated before drill-seeding rye.

Evapotranspiration parameters are required for each crop to estimate the actual evapotranspiration (table 1). The leaf area index (LAI) was measured by an AccuPAR/LAI ceptometer (Decagon Devices, Inc., Pullman, Wash.). The reported values from literature were adopted for the root depth (RD) (Oogathoo, 2006) and crop coefficient (Kc) (Al-Kaisi, 2000). The MIKE SHE recommended values were used for the evapotranspiration parameters for the Kristensen and Jensen model.

### Overland Flow

The overland flow was found to be sensitive to the Manning's number (M), which is the reciprocal of Manning's roughness coefficient (n). A higher value of M leads to faster overland flow and thus lower subsurface

drainage flow. A value of 0.16 was recommended for Manning's n on cropland with a disk tillage system (Engman, 1986), which is equivalent to 6.0 for Manning's M. The Manning's M was subjected to calibration in this study. Considering the small and flat study area, the impact of depression storage depth on surface runoff and drainage flow was relatively small and was set to 0.

### Unsaturated Zone

Undisturbed soil cores at various depths (0-10, 10-20, 20-30, 30-40, 40-60, 60-80, and 80-100 cm) were collected at the modeling plot to determine soil properties, including bulk density, particle size distribution, and saturated hydraulic conductivity (table 2). Other model-required van Genuchten parameters were estimated by the neural network program in the Rosetta model for each soil layer using the measured soil texture and bulk density of each depth (Schaap et al., 2001).

Flow through macropores in unsaturated soil is important for water and solute transport, occurring primarily in wet conditions. Macropore flow was simulated in MIKE SHE using the simple bypass flow method (DHI, 2004). In this method, the actual bypass flow is a function of a user-specified maximum fraction, the minimum water content allowing for bypass flow, and the actual water content of the unsaturated zone. The bypass flow parameters were subjected to calibration.

### Saturated Zone

Vertical and horizontal saturated hydraulic conductivities in the saturated zone substantially influence the characteristics of streamflow and base flow. In the absence of field-measured data, the sensitivity of tile drainage flow to saturated hydraulic conductivities in the saturated zone was investigated (table 3). The depth to the impermeable layer was set to 3.9 m for the study site (Singh et al., 2006).

The subsurface tile drainage system was included as a component of the saturated zone at a depth of 1.06 m below

Table 2. Soil properties of the study site.

Depth (cm)	BD <sup>[a]</sup> (g cm <sup>-3</sup> )	Sand (%)	Silt (%)	K <sub>sat</sub> (cm h <sup>-1</sup> )	θ <sub>r</sub> (cm <sup>-3</sup> cm <sup>-3</sup> )	θ <sub>s</sub> (cm <sup>-3</sup> cm <sup>-3</sup> )
0-10	1.37	32	36	4.8	0.071	0.482
10-20	1.38	32	36	3.3	0.072	0.476
20-30	1.39	33	53	5.1	0.079	0.473
30-40	1.39	40	30	4.1	0.072	0.474
40-60	1.39	46	30	4.1	0.065	0.474
60-90	1.45	44	34	2.6	0.034	0.450
>90	1.46	44	34	2.6	0.033	0.450

[a] BD – bulk density, K<sub>sat</sub> – saturated hydraulic conductivity, θ<sub>r</sub> – residue water content, θ<sub>s</sub> – saturated water content

Table 1. Evapotranspiration parameters for the simulated crops.

Corn				Soybean				Pasture			
Growth Day	LAI <sup>[a]</sup>	RD (mm)	Kc	Growth Day	LAI	RD (mm)	Kc	Growth Day	LAI	RD (mm)	Kc
1	0	0	0.30	1	0	0	0.2	0	1.0	300	0.1
34	0.72	100	0.45	20	1.00	100	0.4	90	1.0	300	0.3
78	3.45	900	0.83	55	6.22	800	0.9	100	1.0	450	0.6
132	3.45	900	0.55	115	6.22	800	1.0	120	2.5	600	0.7
176	2.5	900	0.55	143	4.22	800	0.3	150	4.0	700	0.9
								270	4.0	700	0.9
								330	2.0	300	0.7
								365	1.0	300	0.1

[a] LAI – leaf area index, RD – root depth, Kc – crop coefficient.

**Table 3. Parameters calibration for tile drainage flow.**

	Calibration Range	Final Value
Manning's number (m <sup>1/3</sup> s <sup>-1</sup> )	4.0 – 12.0	6.0
Horizontal hydraulic conductivity (m s <sup>-1</sup> )	1 × 10 <sup>-6</sup> – 1 × 10 <sup>-3</sup>	7.3 × 10 <sup>-6</sup>
Vertical hydraulic conductivity (m s <sup>-1</sup> )	1 × 10 <sup>-6</sup> – 1 × 10 <sup>-3</sup>	7.3 × 10 <sup>-6</sup>
Drainage time constant (h)	0.56 – 2784	5.6
Macropore flow		
Maximum bypass fraction	0.30 – 0.80	0.50
Water content for reduced bypass flow (cm <sup>3</sup> cm <sup>-3</sup> )	0.25 – 0.45	0.30
Minimal water content for bypass flow (cm <sup>3</sup> cm <sup>-3</sup> )	0.15 – 0.35	0.20

the ground surface. In this model, drain line was set in the center of the plot and a grid code map was created to link the drain flow producing cells to the drain line. So, in the model constructed for these plots the actual drain spacing was used although the center grid that was drained is 2 m wide so wider than the drain. In MIKE SHE, subsurface drainage flow occurs when the water table rises above the elevation of tile drains (fig. 1). The drainage flow rate is computed as a linear reservoir and is proportional to the height of the water table above the drainage depth and the specific drainage time constant:

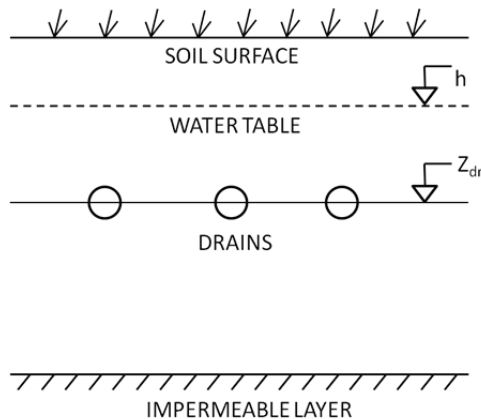
$$Q = \frac{(h - Z_{dr})}{C_{dr}} \quad (1)$$

where Q (m s<sup>-1</sup>) is the drainage flow rate, h (m) and Z<sub>dr</sub> (m) are the levels of water table and tile drain, respectively, and C<sub>dr</sub> (s) is the drainage time constant.

The drainage time constant characterizes the density of the drainage system and the permeability conditions around the drains, and therefore determines the velocity of the drainage. The typical value of the time constant is believed to be between 10 and 120 days (DHI, 2004). However, MIKE SHE is often applied to a much larger watershed size than in this study and no previous calibration has been conducted against observed drainage discharges. Thus, the drainage time constant was subject to calibration for the study plot.

**MODEL CALIBRATION**

The years 2004 and 2005 were used as a “warm up period” to establish initial conditions for MIKE SHE. The



**Figure 1. Schematic representation of the drainage flow computations (Modified from DHI, 2004).**

split-sample calibration-validation method was employed for model calibration and validation. The MIKE SHE model was calibrated against the observed daily tile drainage flow from the monitored plot in 2006 and 2007. Model validation occurred during the following two years (2008 and 2009). To avoid over-parameterization, the number of parameters subjected to adjustment for a distributed model like MIKE SHE should be as small as possible (Regsgaard and Storm, 1995). Therefore, the parameters with measured values such as soil properties and climate data were not calibrated in this study, although subsurface flow may be very sensitive to some of these parameters.

The model performance in simulating subsurface drainage flow was evaluated by both visual comparison of the predicted versus observed hydrographs and some quantitative measures. Two statistical measures were used for quantitative assessment: the Nash-Sutcliffe (NS) model efficiency coefficient (Nash and Sutcliffe, 1970) and coefficient of determination (r<sup>2</sup>).

$$NS = \frac{\sum(O_i - \bar{O})^2 - \sum(P_i - O_i)^2}{\sum(O_i - \bar{O})^2} \quad (2)$$

$$r^2 = \frac{[\sum(O_i - \bar{O})(P_i - \bar{P})]^2}{\sum(O_i - \bar{O})^2 \sum(P_i - \bar{P})^2} \quad (3)$$

where O<sub>i</sub> and P<sub>i</sub> are the observed and predicted values, respectively, and  $\bar{O}$  and  $\bar{P}$  are the mean observed and predicted values, respectively. It should be noted that the days with zero drainage (observed and predicted) were not included in determining the values of NS and r<sup>2</sup>.

The NS value varies from -∞ to 1, measuring the goodness-of-fit between the predicted and observed values. The greater the NS value, the better the model's performance. A negative NS value indicates that even simply using the observation mean would be better than the predicted values by the model. The value of r<sup>2</sup> was used to validate the best-fit line between the predicted and observed values.

**MANAGEMENT SCENARIO SIMULATION**

Strategic integration of land management practices to conventional row-crop agriculture can enhance long-term soil and water quality. Converting row-crops to perennial vegetation has many beneficial effects including reducing soil erosion and water pollution, while minimizing soil disturbance and chemical use. For extensively tile-drained areas, placing tile drains at shallower depths could minimize subsurface drainage and associated pollutant loss while still maximizing crop production (Sands et al., 2003). These two management scenarios were simulated using the calibrated MIKE SHE model: (1) converting row-crops to pasture, and (2) adopting shallow drainage with a higher drain level of 0.75 m. The crop parameters for pasture including LAI and Kc were adopted from FAO (Allen et al., 1998).



## RESULTS AND DISCUSSION

The climate of central Iowa showed a large variability during the study years. The total precipitation at the study site in 2006 was 626 mm representing a “dry” year well below the 30-year (1971-2000) normal annual precipitation of 821 mm in this area. In 2007 and 2008 the total precipitation was 1050 and 926 mm, respectively (table 4; fig. 2), representing “wet” years. The total precipitation in 2009 was 776 mm, which is closer to the long-term normal.

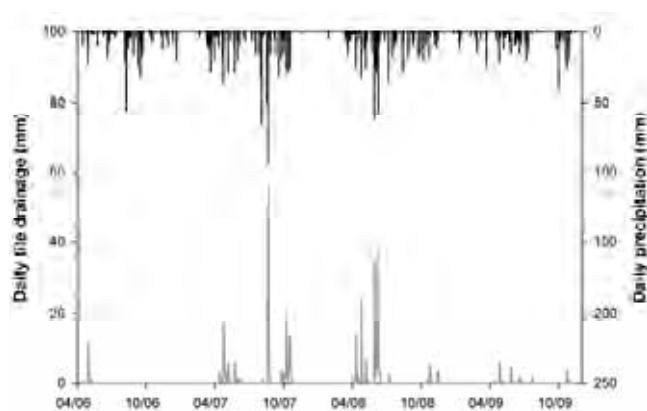
The long-term (1990-2009) average annual drainage during the drainage season (March-November) is 282 mm for this site. Because of the excessive precipitation, the total observed drainage in 2007 and 2008 was 394 and 310 mm, respectively, accounting for about 81% of the total drainage for the entire study period (2006-2009). Most of the drainage occurred in the spring and early summer (April-June) and the fall (September-November) (fig. 2). The wet conditions in the fall of 2007 and the intense precipitation in 2008 led to significant drainage in the spring and early summer of 2008 (table 4). Much less drainage was observed in 2006 and 2009, with only 30 and 134 mm during the monitored period in 2006 and 2009, respectively.

### MODEL CALIBRATION

Model calibration showed that the subsurface tile drainage discharge simulated by MIKE SHE was insensitive to the Manning’s number and thereby the recommended value ( $M = 6.0$ ) for a disked crop system was adopted (table 3). This is not surprising considering the small size and low slope of the plot. In such field conditions, overland flow plays a small role compared to subsurface flow and other hydrologic processes; the predicted surface runoff was only 1.2 and 2.3 mm in 2006 and 2007, respectively. Neverthe-

**Table 4. Annual precipitation and tile drainage flow during April-November in 2006-2009 for study plot.**

Year	Precipitation (mm)	Measured Flow (mm)	Predicted Flow (mm)
2006	626	30	37
2007	1050	394	388
2008	926	305	328
2009	776	136	205
Total	3378	865	956



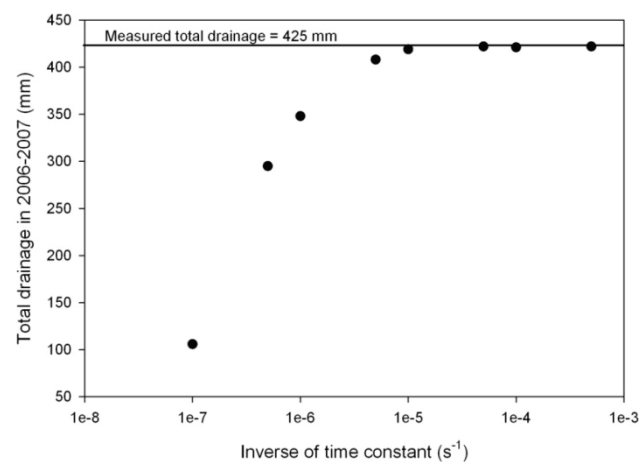
**Figure 2. Precipitation and observed daily tile drainage from April to October during 2006-2009.**

less, the finding that subsurface drainage was insensitive to the Manning’s number is consistent with the findings of Sahoo et al. (2006); they found that the Manning’s number had more influence on the predicted surface flow than base flow. Changes in horizontal and vertical hydraulic conductivities in the saturated zone also showed little impact on the simulated subsurface drainage. The insensitivity of flow simulation to the hydraulic conductivity could be due to the small plot scale of this study. The effect of soil hydraulic properties has been lumped into the drainage time constant at this scale. Therefore, the value of measured saturated hydraulic conductivity ( $7.3 \times 10^{-6} \text{ m s}^{-1}$ ) of the soil profile from 80-100 cm was used for both the horizontal and vertical hydraulic conductivities in the saturated zone (table 3).

The predicted total drainage discharge during the calibration period was very sensitive to the drainage time constant, while inclusion of macropore flows had greater influence on flow peaks than the total drainage. The impacts of these parameters are described in detail below.

### Drainage Time Constant

The drainage time constant ( $C_{dr}$ ) was found to be the most critical parameter when simulating subsurface drainage flow. MIKE SHE lumps the effects of drainage intensity and soil hydraulic properties in drainage time constant, which reflects how fast subsurface water enters into the subsurface drainage system. The influence of the time constant was investigated for values between 0.56 h and 116 days (table 3). As expected, the predicted tile drainage was very sensitive to this parameter, decreasing as the value of  $C_{dr}$  increased until it was greater than 56 h (fig. 3). When  $C_{dr} < 56 \text{ h}$ , the predicted total drainage in 2006-2007 kept nearly the same; however, the model had a better performance for predicting individual storm events for a lower value of  $C_{dr}$ . Generally, the daily tile drainage hydrographs had lower peaks but longer tails for larger values of  $C_{dr}$ . During the calibration period, both the visual inspection of the drainage hydrographs and the statistical measures suggested that the model had the best performance with a value of  $C_{dr}$  equal to 5.6 h.



**Figure 3. Impact of the inverse of drainage time constant on drainage flow. The measured total drainage from April to October during 2006-2007 was 425 mm.**

The time constant value determined from this study was well outside the normal range of 10 to 120 days (DHI, 2004). Several causes may account for the smaller time constant in this study. First, the time constant is usually calibrated against surface discharge at the watershed/basin outlet. Subsurface flow after rain storms could move into the groundwater system as baseflow before it recharges to surface flow, and therefore the influence of the drainage time constant on subsurface drainage or interflow may take a longer time to be detected from monitoring surface discharge. Second, model calibrations were usually performed at a watershed or basin-scale rather than the field-scale employed in this study. For example, a time constant of 33 days was obtained for the 440 km<sup>2</sup> Karup catchment (Refsgaard, 1997), which is about 10<sup>6</sup> times greater than the plot size of this study. As a result, a longer travel time of subsurface drainage flowing out of a watershed would be expected for larger watersheds than smaller watersheds. It should be pointed out that in a large watershed usually only a portion of the entire watershed is artificially drained. This may also contribute to a greater time constant in large watersheds than the one obtained in this study.

### Macropore Flow

When excluding the macropore flow component from the simulation, the total drainage flow in 2006-2007 was estimated by MIKE SHE to be 422 mm, very close to the measured total flow of 425 mm. However, the peak flows after the large storms were generally underestimated, as illustrated in the daily hydrograph of 2007 (fig. 4). Preferential flows through soil macropores and other preferential pathways facilitate subsurface flow when soils are wet. While the addition of macropore flow in the model did not significantly increase the total predicted drainage (424 mm), it improved the prediction on peak flows by providing shortcut pathways for the movement of surface water to the water table. Al-Khudhairy et al. (1999) also observed an improved fit between measured and simulated discharge during wet conditions after including macropore flow in MIKE SHE. The simulated flow peaks were also sensitive to macropore flow parameters, and optimal values

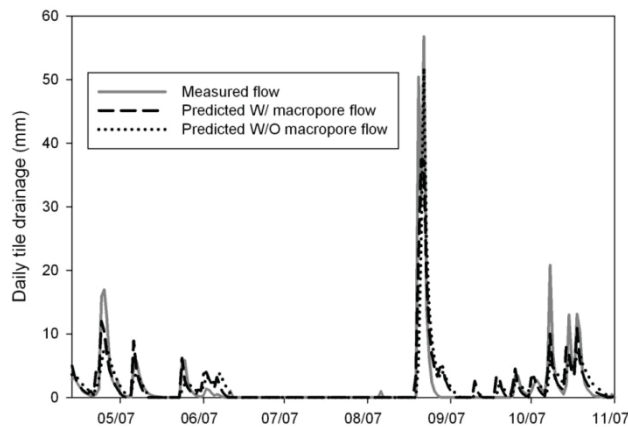


Figure 4. Impact of macropore flow on daily drainage hydrograph in 2007.

shown in table 3 were determined by a combination of visual inspection and quantitative measures. The maximum bypass fraction, water content for reduced bypass flow, and minimal water content for bypass flow were set to 0.6, 0.3, and 0.2, respectively. A larger bypass fraction or a lower water content threshold value enhanced the contribution of macropore flow to the flow peaks. The less successful prediction of peak flow could also be related to the negligence of non-linearities, which execute more impacts when water tables are high, in the subsurface drainage model of MIKE SHE.

### Model Evaluation

Overall, the MIKE SHE model showed a good performance in predicting subsurface tile drainage for the study site, particularly the total drainage and monthly drainage flow (fig. 5). This is especially encouraging considering only two parameters (drainage time constant and macropore flow) were adjusted. The values of NS and  $r^2$  for the monthly drainage prediction were 0.97 and 0.98, respectively. Generally, predictions with model efficiency greater than 0.5 indicate a good model performance (Henriksen et al., 2003). The predicted daily drainage flow also matched the measured flow reasonably well. Overall, the MIKE SHE model showed a good performance during the dry periods and small storm events, but underestimated peak discharges during the large storm events (fig. 5). Despite this, the values of NS and  $r^2$  were 0.78 and 0.80 for

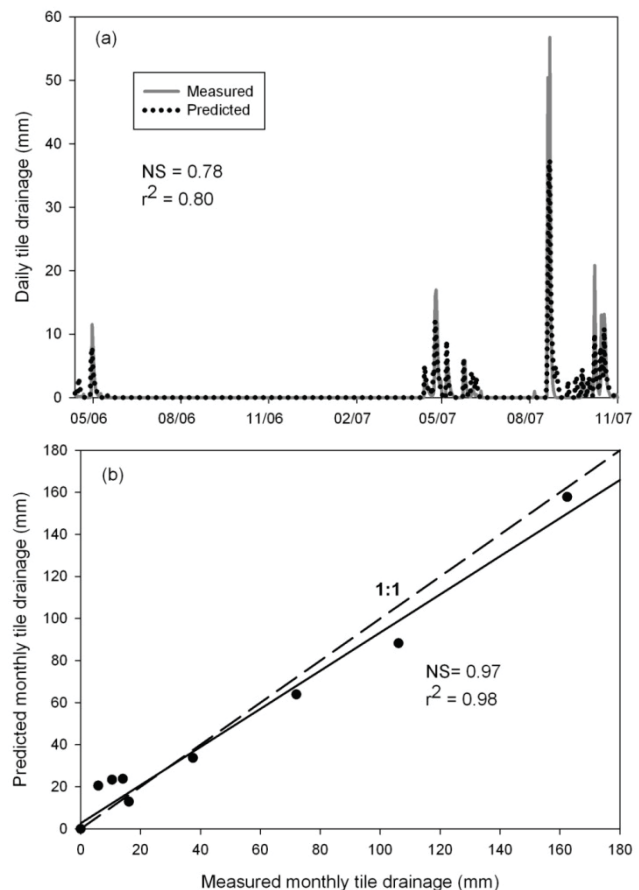


Figure 5. Measured and predicted (a) daily tile drainage and (b) monthly tile drainage during 2006-2007 for model calibration.



the prediction of the daily drainage in 2006-2007, respectively.

### MODEL VALIDATION

The calibrated MIKE SHE model was validated using the daily tile drainage in 2008-2009. Similar to the calibration period, the peaks of large storms in 2008 were somewhat underestimated (fig. 6). However, the model generally overestimated the drainage in 2009, especially for the early summer. In addition, an unobserved peak was also predicted in early April 2008 by the model (fig. 6), possibly because the effect of the winter cover crop on removing excess water during the early spring was underestimated in the model. Nevertheless, the calibrated model had a NS value of 0.73 and  $r^2$  of 0.74 for daily drainage prediction and a NS value of 0.76 and  $r^2$  of 0.79 for monthly drainage prediction, showing a good performance during the validation period. As in the calibration period, the predicted surface runoff was only 0.5 and 0.2 mm in 2008 and 2009, respectively, due to the flat and rough ground surface of the study site.

### MANAGEMENT SCENARIOS SIMULATION

#### Row-Crops vs. Pasture

Land use change from row-crops with cover crops to pasture generally reduced subsurface drainage during the simulation period. The increased evapotranspiration predicted by the model may at least partly account for the subsurface drainage reduction (fig. 7). The effect of land use change on drainage was most evident in 2007, which was a wet year, where tile drainage was reduced by 330 mm (table 5). The simulated drainage reduction in 2008 was only 47 mm despite the fact that 2008 was also a wet year with a total drainage of 310 mm during April-November. This might be attributed to the different pattern of rainfall distribution and field conditions between 2007 and 2008. In 2008, most of the high-intensity storms occurred in the spring and early summer, during which the pasture grasses may be using less water. In addition, due to the wet fall and winter of 2007 and the continuous large storms in the spring of 2008, the lasting wet and cold field conditions may minimize the effect of grasses in removing

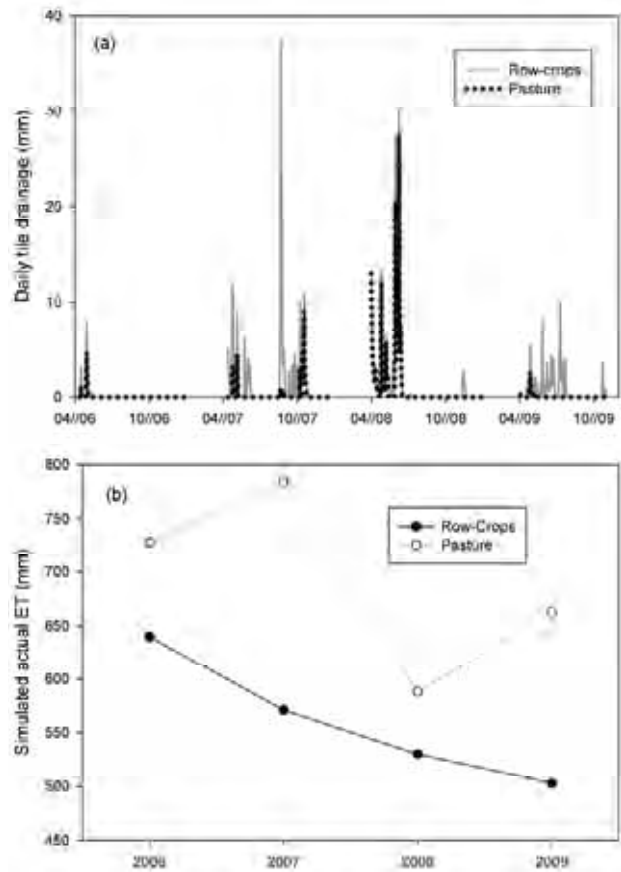


Figure 7. Simulated (a) daily tile drainage and (b) annual actual ET under row-crops vs. pasture.

excess precipitation through transpiration. Consequently, the difference of the simulated ET between row-crops and pasture was only 58 mm in 2008, much smaller than the difference in other years (fig. 7b). The reduced tile drainage by land use change was estimated to be 24 mm in the dry year of 2006. In a 6-year study, Randall et al. (1997) found that total tile flow was 1.6 times higher with row crop systems compared to perennial systems.

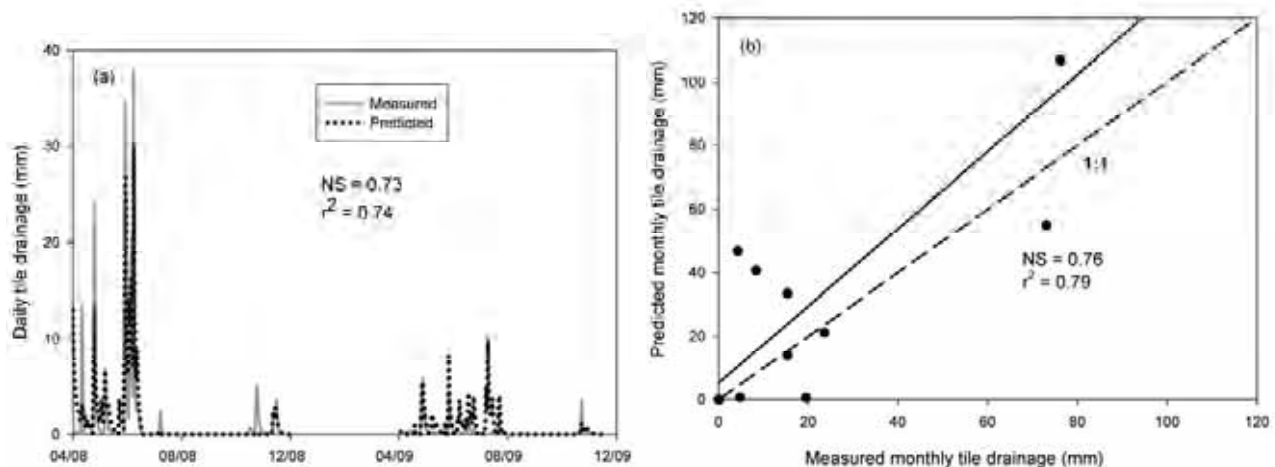


Figure 6. Measured and predicted (a) daily tile drainage and (b) monthly tile drainage during 2008-2009 for model validation.

**Table 5. Predicted tile drainage flow under different management scenarios.**

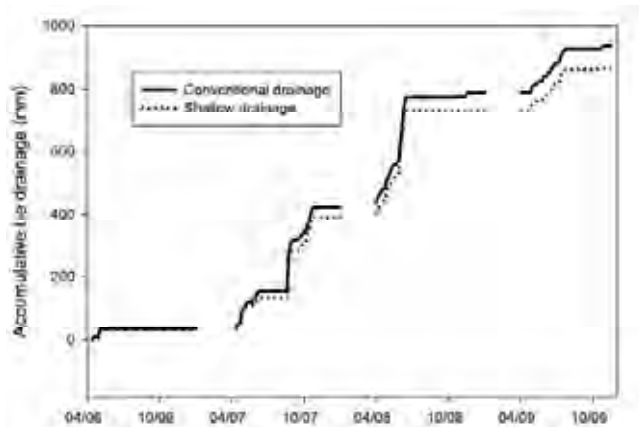
Year	Row-crops with Conventional Drainage (mm)	Row-crops with Shallow Drainage (mm)	Pasture (mm)
2006	37	31	13
2007	388	356	58
2008	365	330	218
2009	205	194	9

### Conventional vs. Shallow Drainage

The simulated tile drainage flow was less sensitive to drainage depth than the drainage time constant (fig. 8), consistent with the findings of other studies (Feyen et al., 2000; Sahoo et al., 2006). As expected, the shallow drainage system at a drain depth of 0.75 m drained less water as compared to the conventional drain depth of 1.05 m. The total subsurface drainage during 2006-2009 was estimated to be 911 mm for shallow drainage, an 8.4% reduction from the conventional drainage system (table 5). The limited reduction may be attributed to the narrow drain spacing of the study site, which is narrower than would be expected in most field conditions. From DRAINMOD simulations at the same location, Singh et al. (2006) found that the subsurface drainage could be reduced by 15% with a shallow drainage system while maintaining the same drainage intensity. Seepage to and from the groundwater is simulated in the MIKE SHE model. Other research showed that the amount of decrease in subsurface drainage by raising water table was impacted by seepage (Skaggs et al., 2010).

## SUMMARY AND CONCLUSIONS

This study attempted to apply the spatially-distributed, physically-based MIKE SHE model for a subsurface tile drainage study in the Midwestern United States. The model was calibrated and validated to subsurface tile drainage flow in a 15.2 × 38 m row-cropped plot in north central Iowa. Among the tested parameters, the drainage time constant and macropore flow had the most notable effect on the predicted drainage. The calibrated drainage time constant was about 5.6 h, which is much smaller than the reported values for



**Figure 8. Impact of tile level on tile drainage flow. For shallow drainage, tile drains were at 0.75 m below ground surface.**

large watersheds. The shorter response time of subsurface flow at the study scale could be responsible for the small drainage time constant in this study. The simulation results suggest that incorporation of macropore flow in hydrological modeling is important for characterizing peak discharges in subsurface drainage systems.

Both the visual comparison of hydrographs and statistical measures (Nash-Sutcliffe model efficiency coefficient and coefficient of determination) indicated that MIKE SHE showed a satisfactory performance in simulating tile drainage flow of the study site during the calibration and validation periods. This is encouraging since only two parameters were adjusted for the calibrated model. The results of modeling changes in land management indicated the potential to reduce subsurface drainage by converting from row-crop to pasture or by shallow drain placement. The role of tile drainage in flood generation and pollutant transport remains under controversy and is difficult to monitor on watershed scales. This study suggests that MIKE SHE has the potential for accurately predicting subsurface flows and ultimately being used in management decisions. The application of MIKE SHE in simulating tile drainage flow from agricultural lands in the Midwestern United States is being investigated at larger scales (e.g., catchment/watershed scales).

## REFERENCES

- Al-Kaisi, M. 2000. Crop water use or evapotranspiration. *Integrated Crop Management* 484, 85-86. Ames, Iowa: Iowa State University-University Extension.
- Al-Khudhairy, D. H. A., J. R. Thompson, H. Gavin, and N. A. S. Hamm. 1999. Hydrological modeling of a drained grazing marsh under agricultural land use and the simulation of restoration management scenario. *Hydrological Science J.* 44(6): 943-971.
- Allen, R. G., L. S. Pereira, D. Raes, and M. Smith. 1998. Crop evapotranspiration: guidelines for computing crop water requirements. Rome: FAO-Irrigation and Drainage.
- Baker, J. L., S. W. Melvin, D. W. Lemke, P. A. Lawlor, W. G. Crumpton, and M. J. Helmers. 2004. Subsurface drainage in Iowa and the water quality benefits and problem. *Drainage VIII Proceedings of the Eighth International Symposium*. St. Joseph, Mich.: ASAE.
- Bengtson, R. L., C. E. Carter, J. L. Fouss, L. M. Southwick, and G. H. Willis. 1995. Agricultural drainage and water quality in Mississippi Delta. *J. of Irrigation and Drainage Engineering* 121(4): 292-295.
- Cambardella, C. A., T. B. Moorman, D. B. Jaynes, T. B. Parkin, and D. L. Karlen. 1999. Water quality in Walnut Creek watershed: NO<sub>3</sub>-N nitrogen in soils, subsurface drainage water and shallow groundwater. *J. of Environmental Quality* 28(1): 25-34.
- Demetriou, C., and J. F. Punthakey. 1999. Evaluating sustainable groundwater management options using the MIKE SHE integrated hydrogeological modelling package. *Environmental Modelling and Software* 14(2-3): 129-140.
- DHI. 2004. *MIKE SHE User Manual*. Hørsholm, Denmark: Danish Hydraulic Institute.
- Engman, E. T. 1986. Roughness coefficients for routing surface runoff. *J. of Irrigation and Drainage Engineering* 112(1): 39-53.
- Feyen, L., R. Vazquez, K. Christiaens, O. Sels, and J. Feyen. 2000. Application of a distributed physically-based hydrological model to a medium size catchment. *Hydrology and Earth System Sciences* 4(1): 47-63.

- Gardenas, A. I., J. Simunek, N. Jarvis, and M.T. van Genuchten. 2006. Two-dimensional modeling of preferential water flow and pesticide transport from a tile-drained field. *J. of Hydrology* 329(3-4): 647-660.
- Gardner, W. K., M. F. Drendel, and G. K. McDonald. 1994. Effects of subsurface drainage, cultivation and stubble retention on soil porosity and crop growth in a high rainfall area. *Australian J. of Experimental Agriculture* 34(3): 411-418.
- Helmers, M. J., D. E. Eisenhauer, T. G. Franti, and M. G. Dosskey. 2005. Modeling sediment trapping in a vegetative filter accounting for converging overland flow. *Trans. ASAE* 48(2): 541-555.
- Henriksen, H. J., L. Trolborg, P. Nyegaard, T. O. Sonnenborg, J. C. Refsgaard, and B. Madsen. 2003. Methodology for construction, calibration and validation of a national hydrological model for Denmark. *J. of Hydrology* 280(1-4): 52-71.
- Irwin, R. W., and H. R. Whiteley. 1983. Effects of land drainage on stream flow. *Canadian Water Resources J.* 8(2): 88-103.
- Jaber, F. H., and S. Shukla. 2005. Hydrodynamic modeling approaches for agricultural storm water impoundments. *J. of Irrigation and Drainage Eng.* 131(4): 307-315.
- Kanwar, R. S., R. M. Cruse, M. Ghaffarzadeh, A. Bakhsh, D. L. Karlen, and T. B. Bailey. 2005. Corn soybean and alternative cropping systems effected on NO<sub>3</sub>-N leaching losses in subsurface drainage water. *Applied Eng. in Agric.* 21(2): 181-188.
- Karl, T. R., J. M. Melillo, and T. C. Peterson. 2009. Global climate change impacts in the United States. New York, N.Y.: Cambridge University Press.
- Konyha, K. D., R. W. Skaggs, and J. W. Gilliam. 1992. Effects of drainage and water management practices on hydrology. *J. of Irrigation and Drainage Eng.* 118(5): 807-819.
- Kristensen, K. J., and S. E. Jensen. 1975. A model for estimating actual evapotranspiration from potential evapotranspiration. *Nordic Hydrology* 6(3): 170-188.
- Lawlor, P. A., M. J. Helmers, J. L. Baker, S. W. Melvin, and D. W. Lemke. 2008. Nitrogen application rate effect on NO<sub>3</sub>-N-nitrogen concentration and loss in subsurface drainage for a corn-soybean rotation. *Trans. ASAE* 51(1): 83-94.
- Nash, J. E., and J. V. Sutcliffe. 1970. River flow forecasting through conceptual models. Part I. A discussion of principles. *J. of Hydrology* 10(3): 282-290.
- Oogathoo, S. 2006. Runoff simulation in the Canagagigue creek watershed watershed using the MIKE SHE Model. M.S. thesis. Sainte Anne-de-Bellevue, QC: McGill University, Department of Agricultural and Biosystems Engineering.
- Rabalais, N. N., R. E. Turner, and W. J. Wiseman, Jr. 2001. Hypoxia in the Gulf of Mexico. *J. of Environmental Quality* 30(2): 320-329.
- Randall, G. W., D. R. Huggins, D. J. Fuchs, W. W. Nelson., and J. L. Anderson. 1997. Nitrate losses through subsurface tile drainage in conservation reserve program, alfalfa, and row crop systems. *J. of Environmental Quality* 26(5): 1240-1247.
- Refsgaard, J. C. 1997. Parameterisation, calibration and validation of distributed hydrological models. *J. of Hydrology* 198(1-4): 69-97.
- Refsgaard, J. C., and B. Storm. 1995. MIKE SHE. In *Computer Models of Watershed Hydrology*, 809-846. V. Singh, ed. Highlands Ranch, Colo.: Water Resources Publications.
- Sahoo, G. B., C. Ray, and E. H. De Carlo. 2006. Calibration and validation of a physically distributed hydrological model, MIKE SHE, to predict streamflow at high frequency in a flashy mountainous Hawaii Stream. *J. of Hydrology* 327(1-2): 94-109.
- Sands, G. R., L. M. Busman, W. E. Ruger, and B. Hansen. 2003. The impact of drain depth on water quality in a cold climate. ASAE Paper No. 032365. St. Joseph, Mich.: ASAE.
- Schaap, M. G., F. J. Leij, and M. Th. van Genuchten. 2001. Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedo-transfer functions. *J. of Hydrology* 251(3): 163-176.
- Shipitalo, M. J., V. Nuutinen, and K. R. Butt. 2004. Interaction of earthworm burrows and cracks in a clayey, subsurface drained, soil. *Applied Soil Ecology* 26(3): 209-217.
- Singh, R., J. C. Refsgaard, L. Yde, G. H. Jorgensen, and M. Thorsen. 1997. Hydraulic-hydrological simulations of canal-command for irrigation water management. *Irrigation and Drainage Systems* 11(3?): 185-213.
- Singh, R., M. J. Helmers, and Z. Qi. 2006. Calibration and validation of DRAINMOD to design subsurface drainage systems for Iowa's tile landscape. *Agricultural Water Management* 85(3): 221-232.
- Skaggs, R. W., and R. G. Broadhead. 1982. Drainage strategies and peak flood flows. ASAE Paper No. 822054. St. Joseph, Mich.: ASAE.
- Skaggs, R. W., M. A. Breve, and J. W. Gilliam. 1994. Hydrologic and water quality impact of agricultural drainage. *Critical Reviews in Environmental Science and Technology* 24(1): 1-32.
- Skaggs, R. W., M. A. Breve, and J. W. Gilliam. 1995. Simulation of drainage water quality with DRAINMOD. *Irrigation and Drainage Systems* 9(3): 259-277.
- Skaggs, R. W., M. A. Youself, J. W. Gilliam, and R. O. Evans. 2010. Effect of controlled drainage on water and nutrient balances in drained lands. *Trans. ASABE* 53(6): 1843-1850.
- Thompson, J. R., H. R. Sorenson, H. Gavin, and A. Refsgaard. 2004. Application of the coupled MIKE SHE/MIKE 11 modelling system to a lowland wet grassland in southeast England. *J. of Hydrology* 293(1-4): 151-179.
- Wiskow, W. and R. R. van der Ploeg. 2003. Calculation of drain spacings for optimal rainstorm flood control. *J. of Hydrology* 272(1-4): 163-174.