Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models

Mukesh K. Tiwari1 and Jan Adamowski2

Received 26 July 2012; revised 20 August 2013; accepted 4 September 2013; published 9 October 2013.

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 naive 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.

1 Department of Soil and Water Engineering, College of Agricultural and Technology, Anand Agricultural University, Godhra, Gujarat, India.
2 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)

©2013. American Geophysical Union. All Rights Reserved. 0043-1397/13/10.1002/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 et al., 2005].
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 nonlinearity, 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. However, 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. This property of wavelet analysis is different from Fourier analysis. Wavelet analysis is different from Fourier analysis. 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 number of scale-variable processes. Because of its localization properties in time and frequency, wavelet analysis is well suited for analyzing signals that are non-stationary or non-linear. The wavelet transform can be denoted as:

\[ \psi(t) dt = 0. \]  
(2)

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

\[ \int_{-\infty}^{+\infty} \psi(t) dt = 0. \]  
(2)

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.

[12] The successive wavelet, can be derived as [Kisi, 2010]

\[ \psi_{a,b}(t) = |a|^{-\frac{1}{2}} \psi \left( \frac{t-b}{a} \right) \]  
(3)

where \( m \) and \( n \) are integers that determine the magnitude of wavelet dilation and translation, respectively, \( d_0 \) is a specified dilation step greater than 1 (most commonly \( d_0 = 2 \)), and \( b_0 \) is the location parameter which must be greater than zero (most commonly \( b_0 = 1 \)). For a discrete time series \( f(t) \), assuming \( d_0 = 2 \) and \( b_0 = 1 \), the DWT simplifies as [Kisi, 2010]

\[ W_f(m, n) = 2^{-m/2} \sum_{i=0}^{N-1} f(t) \psi \left( 2^{-m} i - n \right) \]  
(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 \( (t = 0, 1, 2, \ldots, N - 1) \), 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 \( 0 < n < 2^M - 1 \), 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 \( 1 + 2 + 4 + 8 + \ldots + 2M - 1 = N - 1 \). Moreover, the term \( \mathcal{W} \) 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]

\[
f(t) = \mathcal{W} + \sum_{m=1}^{M} \sum_{n=0}^{2^{m-1}} W_f(m,n)2^{-\frac{m}{2}} \psi(t - n)
\]

or further simplified as [Nourani et al., 2009]

\[
f(t) = \mathcal{W}(t) + \sum_{m=1}^{M} W_m(t)
\]

where \( \mathcal{W}(t) \) is the approximation subsignal at level \( M \), and \( W_m(t) \) is the detailed subsignal at each level \( m = 1, 2, \ldots, M \).

### 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 variable with \( y_i \), the corresponding output variable, and \( n \) number of samples are drawn from \( F \), resulting in a random data set sample, denoted as \( T_n = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \). The empirical distribution function for \( T_n \) is \( F \) with a mass of 1/\( n \) for each \( t_1, t_2, \ldots, t_n \). Similar to sampling from the unknown distribution \( F \), an \( n \) number of samples of \( t_i = (x_i, y_i) \) are taken from \( F \) that are i.i.d and then replaced. One bootstrap realization would be the resulting random sample set \( T' = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \), and a set of bootstrap samples of \( T', T', \ldots, T', \ldots, T' \) can be produced. The total number of bootstrap samples, \( S \), usually ranges from 50 to 200 samples [Efron, 1979]. For each \( T' \), a NN model is developed and trained using all \( n \) observations. The NN output, \( f_{NN}(x_i, w_i/T') \), 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 \( T' \). 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]

\[
E_0 = \frac{\sum_{i=1}^{S} \sum_{a=1}^{A} (y_i - f_{NN}(x_i, w_i/T'))^2}{\sum_{i=1}^{S} (A)}
\]

where \( f_{NN}(x_i, w_i/T') \) is, again, the output of the NN developed from the bootstrap sample \( T' \), in which \( x_i \) is a particular input vector and \( w_i \) 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 [Jiao and Calver, 2006]

\[
\hat{y}(x) = \frac{1}{S} \sum_{s=1}^{S} f_{NN}(x, w_s/T')
\]

and the variance is given by

\[
\sigma^2(x) = \frac{\sum_{s=1}^{S} \sum_{l=1}^{A} (y_l - f_{NN}(x, w_s/T'))^2}{S - 1}
\]

[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 \( \% \) confidence limits. 100 \( \times \) (1 - (1 - \( \alpha \))) CI covering the mean/ensemble demand \( \hat{y}(x) \) can be estimated using the following equation [Efron and Tibshirani, 1993]

\[
CI = [UB, LB] = \left[ \hat{y}(x) + \frac{T_{\alpha/2}}{\sigma(x)}, \hat{y}(x) - \frac{T_{\alpha/2}}{\sigma(x)} \right]
\]

where UB is the upper band, LB is the lower band, \( \sigma(x) \) is the standard deviation of \( S \) bootstrapped estimates, \( T_{\alpha/2} \) 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 Baillets, 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.

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

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

All the methodological issues need to be considered for the development of a robust NN model [Maier et al.,...
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

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Period</th>
<th>Number of Data Patterns for Daily and Weekly Water Demand Forecasting</th>
<th>Number of Data Patterns for Monthly Water Demand Forecasting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>2/27/1999 to 12/31/2007</td>
<td>3230</td>
<td>107</td>
</tr>
<tr>
<td>Cross-validation</td>
<td>1/1/2008 to 12/31/2008</td>
<td>366</td>
<td>13</td>
</tr>
<tr>
<td>Testing</td>
<td>1/1/2009 to 8/6/2010</td>
<td>583</td>
<td>18</td>
</tr>
</tbody>
</table>
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 square 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 ($P_d$), 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].
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 information 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 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.
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

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
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.

### 5.1.4. Daily Water Demand Forecasting Using WBNN Models

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

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

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

<table>
<thead>
<tr>
<th>Lead Time</th>
<th>Best Model Structure</th>
<th>Hidden Neurons (HN)</th>
<th>Performance Indices</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 day</td>
<td>NN</td>
<td>13</td>
<td>$R^2$ RMSE (ML/d) Pdv (%) MAE (ML/d) PI</td>
</tr>
<tr>
<td></td>
<td>Same inputs as in NN</td>
<td></td>
<td>0.92 5.88 0.88 4.17 0.45</td>
</tr>
<tr>
<td>3 day</td>
<td>NN</td>
<td>5</td>
<td>0.77 9.92 2.52 7.15 0.41</td>
</tr>
<tr>
<td>5 day</td>
<td>NN</td>
<td>11</td>
<td>0.71 10.91 4.80 7.11 0.37</td>
</tr>
<tr>
<td>1 day</td>
<td>BNN</td>
<td></td>
<td>0.90 6.09 1.66 4.41 0.41</td>
</tr>
<tr>
<td>3 day</td>
<td>BNN</td>
<td></td>
<td>0.76 10.70 2.42 7.96 0.31</td>
</tr>
<tr>
<td>5 day</td>
<td>BNN</td>
<td></td>
<td>0.64 12.55 0.07 9.14 0.07</td>
</tr>
<tr>
<td>1 day</td>
<td>WNN</td>
<td>0.98 3.11 0.57 2.21 0.84</td>
<td></td>
</tr>
<tr>
<td>3 day</td>
<td>WNN</td>
<td>0.94 5.33 0.82 4.17 0.83</td>
<td></td>
</tr>
<tr>
<td>5 day</td>
<td>WNN</td>
<td>0.90 7.81 0.37 6.30 0.68</td>
<td></td>
</tr>
<tr>
<td>1 day</td>
<td>WBNN</td>
<td>0.98 4.09 0.97 2.99 0.73</td>
<td></td>
</tr>
<tr>
<td>3 day</td>
<td>WBNN</td>
<td>0.96 6.08 1.04 4.81 0.78</td>
<td></td>
</tr>
<tr>
<td>5 day</td>
<td>WBNN</td>
<td>0.90 7.29 0.54 5.82 0.72</td>
<td></td>
</tr>
<tr>
<td>1 day</td>
<td>ARIMA</td>
<td>0.90 6.51 0.58 4.85 0.33</td>
<td></td>
</tr>
<tr>
<td>3 day</td>
<td>ARIMA</td>
<td>0.70 10.55 1.32 7.60 0.30</td>
<td></td>
</tr>
<tr>
<td>5 day</td>
<td>ARIMA</td>
<td>0.59 14.57 0.79 11.33 0.12</td>
<td></td>
</tr>
<tr>
<td>1 day</td>
<td>ARIMAX</td>
<td>0.91 6.55 2.70 4.72 0.39</td>
<td></td>
</tr>
<tr>
<td>3 day</td>
<td>ARIMAX</td>
<td>0.71 10.42 1.4 7.44 0.37</td>
<td></td>
</tr>
<tr>
<td>5 day</td>
<td>ARIMAX</td>
<td>0.64 11.44 2.25 8.83 0.41</td>
<td></td>
</tr>
</tbody>
</table>

*Table 2. Correlations Between Different Wavelet Subtime Series and the Original Water Demand Time Series in Montreal

<table>
<thead>
<tr>
<th>Wavelet Subtime Series</th>
<th>Daily</th>
<th>Monthly</th>
</tr>
</thead>
<tbody>
<tr>
<td>A3</td>
<td>0.94</td>
<td>0.55</td>
</tr>
<tr>
<td>D1</td>
<td>0.16</td>
<td>0.20</td>
</tr>
<tr>
<td>D2</td>
<td>0.25</td>
<td>0.04</td>
</tr>
<tr>
<td>D3</td>
<td>0.19</td>
<td>0.65</td>
</tr>
<tr>
<td>Original</td>
<td>1.00</td>
<td>0.64</td>
</tr>
</tbody>
</table>

*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

$^{a}$WatDmand ($t + 1$) = 1 day/week/month lead water demand forecast; WatDmand(t) = Total water demand at time (t).
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 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

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 naive 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

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

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 naive persistence model, there is a need to apply hybrid approaches to improve the performance of NN models for higher step weekly water demand forecasting.

5.2.4. Weekly Water Demand Forecasting Using WBNN Models

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>
<thead>
<tr>
<th>Lead Time</th>
<th>Best Model Structure</th>
<th>HN</th>
<th>$R^2$</th>
<th>RMSE (ML/d)</th>
<th>$P_{res}$ (%)</th>
<th>MAE (ML/d)</th>
<th>PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 week</td>
<td>WatDmand(t), WatDmand(t-6), MaxT(t), MaxT (t-6), TotP(t), TotP(t-6)</td>
<td>NN</td>
<td>5</td>
<td>0.67</td>
<td>11.83</td>
<td>4.07</td>
<td>8.59</td>
</tr>
<tr>
<td>2 week</td>
<td>WatDmand(t), WatDmand(t-6), MaxT(t), MaxT (t-6), TotP(t), TotP(t-6)</td>
<td>8</td>
<td>0.59</td>
<td>17.59</td>
<td>2.09</td>
<td>14.45</td>
<td>0.17</td>
</tr>
<tr>
<td>1 week</td>
<td>Same inputs as in NN</td>
<td>BNN</td>
<td>0.68</td>
<td>12.33</td>
<td>3.15</td>
<td>9.17</td>
<td>0.10</td>
</tr>
<tr>
<td>2 week</td>
<td>Same inputs as in NN</td>
<td>0.57</td>
<td>18.02</td>
<td>2.75</td>
<td>15.22</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>1 week</td>
<td>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</td>
<td>WNN</td>
<td>5</td>
<td>0.76</td>
<td>12.07</td>
<td>-1.76</td>
<td>8.90</td>
</tr>
<tr>
<td>2 week</td>
<td>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</td>
<td>3</td>
<td>0.72</td>
<td>15.49</td>
<td>-0.14</td>
<td>10.93</td>
<td>0.28</td>
</tr>
<tr>
<td>1 week</td>
<td>Same inputs as in WNN</td>
<td>WBNN</td>
<td>0.78</td>
<td>10.06</td>
<td>-0.58</td>
<td>7.49</td>
<td>0.46</td>
</tr>
<tr>
<td>2 week</td>
<td>Same inputs as in WNN</td>
<td>0.68</td>
<td>18.76</td>
<td>1.47</td>
<td>16.28</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>1 week</td>
<td>ARIMA</td>
<td>0.68</td>
<td>11.74</td>
<td>0.03</td>
<td>7.49</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>2 week</td>
<td>ARIMA</td>
<td>0.52</td>
<td>16.95</td>
<td>4.36</td>
<td>14.46</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>1 week</td>
<td>ARIMAX</td>
<td>0.68</td>
<td>11.89</td>
<td>2.52</td>
<td>8.80</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>2 week</td>
<td>ARIMAX</td>
<td>0.51</td>
<td>17.03</td>
<td>3.90</td>
<td>14.58</td>
<td>0.14</td>
<td></td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Lead Time</th>
<th>Best Model Structure</th>
<th>HN</th>
<th>$R^2$</th>
<th>RMSE (ML/month)</th>
<th>$P_{0.05}$ (%)</th>
<th>MAE (ML/month)</th>
<th>PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN 1 month</td>
<td>WatDmand(t), WatDmand(t-11), Maxt(t), Maxt(t-11), TotP(t), TotP(t-11)</td>
<td>2</td>
<td>0.87</td>
<td>22.18</td>
<td>4.51</td>
<td>20.31</td>
<td>0.06</td>
</tr>
<tr>
<td>NN 2 month</td>
<td>WatDmand(t), WatDmand(t-10), Maxt(t), Maxt(t-11), TotP(t), TotP(t-11)</td>
<td>8</td>
<td>0.70</td>
<td>24.27</td>
<td>4.69</td>
<td>22.33</td>
<td>0.54</td>
</tr>
<tr>
<td>BNN 1 month</td>
<td>Same inputs as in NN</td>
<td>0.77</td>
<td>21.74</td>
<td>3.84</td>
<td>19.70</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>BNN 2 month</td>
<td>Same inputs as in NN</td>
<td>0.74</td>
<td>23.03</td>
<td>4.34</td>
<td>20.97</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>WNN 1 month</td>
<td>A3, d1, d2, d3 components of WatDmand(t), A3 and d3 components of Maxt(t) and d3 component of TotP(t)</td>
<td>3</td>
<td>0.96</td>
<td>5.94</td>
<td>−0.33</td>
<td>4.82</td>
<td>0.81</td>
</tr>
<tr>
<td>WNN 2 month</td>
<td>A3, d1, d2, d3 components of WatDmand(t), A3 and d3 components of Maxt(t) and d3 component of TotP(t)</td>
<td>5</td>
<td>0.91</td>
<td>7.21</td>
<td>0.23</td>
<td>5.99</td>
<td>0.89</td>
</tr>
<tr>
<td>WBNN 1 month</td>
<td>Same inputs as in WNN</td>
<td>0.77</td>
<td>11.13</td>
<td>−0.68</td>
<td>8.62</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>WBNN 2 month</td>
<td>Same inputs as in WNN</td>
<td>0.54</td>
<td>13.49</td>
<td>2.25</td>
<td>10.94</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td>ARIMA 1 month</td>
<td></td>
<td>0.57</td>
<td>16.98</td>
<td>4.29</td>
<td>14.86</td>
<td>0.47</td>
<td></td>
</tr>
<tr>
<td>ARIMA 2 month</td>
<td></td>
<td>0.67</td>
<td>19.68</td>
<td>2.99</td>
<td>17.70</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>ARIMAX 1 month</td>
<td></td>
<td>0.90</td>
<td>19.20</td>
<td>−4.06</td>
<td>18.21</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>ARIMAX 2 month</td>
<td></td>
<td>0.90</td>
<td>18.62</td>
<td>−3.94</td>
<td>17.61</td>
<td>0.72</td>
<td></td>
</tr>
</tbody>
</table>
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 datasets 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 days; 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 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.
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

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

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

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

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 for WBNN predicted confidence bands compared to the actual values included in the BNN forecasted confidence bands.

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 forecasts.
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 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.
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.

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.
5.4.4. Performance Comparison of WBNNs With Different Numbers of Bootstrap Resamples

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 the WBNN model developed using 50 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 included 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 shown in Figure 1c, and the lower autocorrelation in the case of monthly water demand time series, as 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.

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

<table>
<thead>
<tr>
<th>Performance Indices</th>
<th>500</th>
<th>200</th>
<th>100</th>
<th>50</th>
<th>25</th>
<th>500</th>
<th>200</th>
<th>100</th>
<th>50</th>
<th>25</th>
<th>500</th>
<th>200</th>
<th>100</th>
<th>50</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.99</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.95</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.91</td>
<td>0.91</td>
<td>0.90</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>RMSE (ML/day)</td>
<td>4.08</td>
<td>4.24</td>
<td>4.09</td>
<td>4.43</td>
<td>3.82</td>
<td>6.64</td>
<td>6.16</td>
<td>6.08</td>
<td>6.57</td>
<td>6.40</td>
<td>8.13</td>
<td>7.75</td>
<td>7.29</td>
<td>7.49</td>
<td>7.48</td>
</tr>
<tr>
<td>$P_{pe}$ (%)</td>
<td>−1.07</td>
<td>−0.92</td>
<td>−0.97</td>
<td>−1.08</td>
<td>−1.72</td>
<td>0.90</td>
<td>1.05</td>
<td>1.04</td>
<td>1.12</td>
<td>1.32</td>
<td>1.09</td>
<td>0.91</td>
<td>0.54</td>
<td>0.26</td>
<td>−0.17</td>
</tr>
<tr>
<td>MAE (ML/day)</td>
<td>3.00</td>
<td>3.13</td>
<td>2.99</td>
<td>3.26</td>
<td>2.76</td>
<td>5.29</td>
<td>4.89</td>
<td>4.81</td>
<td>5.23</td>
<td>5.09</td>
<td>6.63</td>
<td>6.27</td>
<td>5.82</td>
<td>6.01</td>
<td>5.99</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Performance Indices</th>
<th>1 week</th>
<th>2 week</th>
<th>1 month</th>
<th>2 month</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.88</td>
<td>0.74</td>
<td>0.77</td>
<td>0.77</td>
</tr>
<tr>
<td>RMSE (ML/day)</td>
<td>10.31</td>
<td>10.69</td>
<td>10.66</td>
<td>11.85</td>
</tr>
<tr>
<td>$P_{pe}$ (%)</td>
<td>−1.53</td>
<td>−5.97</td>
<td>−0.58</td>
<td>−1.95</td>
</tr>
<tr>
<td>MAE (ML/month)</td>
<td>7.67</td>
<td>9.61</td>
<td>7.49</td>
<td>9.44</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Performance Indices</th>
<th>1 month</th>
<th>2 month</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.86</td>
<td>0.89</td>
</tr>
<tr>
<td>$P_{pe}$ (%)</td>
<td>−4.39</td>
<td>−0.92</td>
</tr>
<tr>
<td>MAE (ML/month)</td>
<td>13.92</td>
<td>7.48</td>
</tr>
</tbody>
</table>

6. Conclusions

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 ($P_{pe}$), mean absolute
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


Environment Canada (2010), Available at http://www.climate.weatheroffice.gc.ca.


6506


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.


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.


