Water demand forecasting using extreme learning machines

Mukesh TIWARI1) ABCDEF, Jan ADAMOWSKI2) ABCDEF, Kazimierz ADAMOWSKI3) A

1) Anand Agricultural University, Department of Soil and Water Engineering, College of Agricultural and Technology, Gujarat, India; e-mail: tiwari.itikgp@gmail.com
2) McGill University, Faculty of Agricultural and Environmental Sciences, Department of Bioresource Engineering, Quebec, Canada, H9X 3V9; e-mail: jan.adamowski@mcgill.ca
3) University of Ottawa, Department of Civil Engineering, Canada; e-mail: kadamowski@hotmail.com


Abstract

The capacity of recently-developed extreme learning machine (ELM) modelling approaches in forecasting daily urban water demand from limited data, alone or in concert with wavelet analysis (W) or bootstrap (B) methods (i.e., ELM, ELMW, ELMB), was assessed, and compared to that of equivalent traditional artificial neural network-based models (i.e., ANN, ANNW, ANNB). The urban water demand forecasting models were developed using 3-year water demand and climate datasets for the city of Calgary, Alberta, Canada. While the hybrid ELMB and ANNb models provided satisfactory 1-day lead-time forecasts of similar accuracy, the ANNw and ELMw models provided greater accuracy, with the ELMW model outperforming the ANNW model. Significant improvement in peak urban water demand prediction was only achieved with the ELMW model. The superiority of the ELMW model over both the ANNw or ANNb models demonstrated the significant role of wavelet transformation in improving the overall performance of the urban water demand model.

Key words: artificial neural networks, bootstrap, Canada, extreme learning machines, uncertainty, water demand forecasting, wavelets

INTRODUCTION

Attributable to demographic expansion and industrial development, the rapid rise in worldwide urban water consumption has placed potable water distribution systems under stress. Given the advent of climate change, these problems will likely become more acute in the future [SAADAT et al. 2011; ARAGHI et al. 2015]. Accurate forecasting of short-term water demand can contribute to the efficient operation and management of urban water supply systems, resulting in demand being met efficiently and sustainably [ADAMOWSKI et al. 2013; CAMPISI-PINTO et al. 2012; TIWARI, ADAMOWSKI 2014]. The estimation of future urban water demand is therefore essential to the sustainable planning of regional water-supply systems [ADAMOWSKI et al. 2013; TIWARI, ADAMOWSKI 2015a, b; ZHOU et al. 2002]. Given increases in the diverse components of urban water demand (e.g., residential, public, industrial and commercial use – HANEMANN [1998]), water stress and scarcity have become critical issues [ADAMOWSKI et al. 2012a, b; 2013; DAVIS, KIEFER 2005; GOYAL et al. 2014; HAI-DARY et al. 2013; KAYAGA, SMOUT 2008]. Furthermore, multiple-scale interactions between individuals and natural systems create a further range of urban water demand management challenges [HOUSE-PETERS, CHANG 2011]. Short-term urban water de-
mand forecasts play a significant role in the optimal operation of pumps, wells, and reservoirs, as well as in informing decisions regarding balanced water resource allocation in the face of urgent water needs [HERRERA et al. 2010; JAIN, ORMSBEE 2002; KAME'ENUI 2003]. Urban water is generally allocated according to the experience of operators and average water demand; however, accurate and reliable forecasts of short-term demand can help operators provide water in a more efficient and sustainable manner [ZHOU et al. 2002].

This study focuses on fast, efficient methods for short-term (1-day lead time) urban water demand forecasts, in an effort to achieve accurate and reliable forecasts of water demand for the City of Calgary, Alberta, Canada. Traditionally, linear regression, trend-extrapolation, and time-series techniques have been applied in forecasting water resources operations variables, particularly in the domain of urban water demand [ADAMOWSKI et al. 2012a; BELAYNEH et al. 2014]. Short-term water demand data generally exhibits nonlinear and nonstationary behaviour [GHIASSI et al. 2008] at multiple spatial and temporal scales [HOUSE-PETERS, CHANG 2011]. Non-stationarity, such as that attributable to seasonal variations and trends, significantly lowers modelling accuracy for time series, generally leading to poor predictions in operational applications [ADAMOWSKI et al. 2009; 2010; FRANCESCO, BERND 2000; NALLEY et al. 2012; 2013; PINGALE et al. 2014; RATHINASAMY et al. 2013; 2015]. Wavelet transformation, a time–frequency representation of a signal present at many different intervals in the time domain, provides considerable information about the physical structure of the time series data [BELAYNEH et al. 2014; DAUBECHIES 1990; KARRAN et al. 2014; NOURANI et al. 2014]. Wavelet analysis uses a mother wavelet function to decompose non-stationary data into multiple scale-specific time series [NALLEY et al. 2012; 2013] and thereby helps to distinguish among daily, weekly, and seasonal cycles inherent in water demand.

Wavelet-transformation-based artificial neural networks (ANNw) have been found to be more accurate than multiple linear regression, time-series or regular artificial neural network (ANN) models in forecasting regional drought [KIM, VALDES 2003], rainfall–runoff [ANCTIL, TAPE 2004; NOURANI et al. 2009], monthly and daily streamflow [ADAMOWSKI 2007; ADAMOWSKI, SUN 2010; KISI 2008; 2009; MAHESWARAN, KHOSA 2012; NAYAK et al. 2013; TIWARI et al. 2013], monthly groundwater levels [ADAMOWSKI, CHAN 2011], and short-term urban water demand [ADAMOWSKI et al. 2012a, b; c; TIWARI, ADAMOWSKI 2013]. Other hybrid methods proposed in the hydrological forecasting literature are the bootstrap-based ANN (ANNb) [TIWARI, CHATTERJEE 2010a, b], fuzzy neural networks [ALVISI, FRANCHINI 2011], and grey neural networks [ALVISI, FRANCHINI 2012]. Besides reducing uncertainty in the variance by mimicking randomness [EFRON, TIBSHIRANI 1993], ANNb models are simpler and easier to use in addressing uncertainty in an operational setting compared to Bayesian approaches [ISUKAPALLI, GEORGOPOULOS 2001]. Several studies have shown ANNw models to outperform standard ANN models [ABRAHART 2003; HAN et al. 2007; JEONG, KIM 2005; JIA, CULVER 2006; SHARMA, TIWARI 2009; SRIVASTAV et al. 2007; TIWARI, CHATTERJEE 2010a]. Both ANNw and ANNb hybrid approaches can be combined to form a wavelet-bootstrap-ANN (ANNwbb) model with the potential ability to achieve greater accuracy and reliability in real time water demand forecasting. However, this has not been undertaken to date.

In the present study a wavelet-extreme learning machine (ELMw) [HUANG et al. 2006; 2015] based water demand model, developed with limited data (limited years of dataset), is proposed. The ELM is a fast three-step model designed to use a Single Layer Feedforward Neural Network with hidden neurons and randomly chosen weights. The hidden layer learns patterns from distinct observations and therefore requires no parameter tuning, only a predefined network. The ELM is free from the complications faced by gradient-based algorithms (e.g., learning rate, learning epochs and local minima) [ACHARYA et al. 2014; BELAYNEH, ADAMOWSKI 2014; ŞAHIN et al. 2014]. Despite their widespread use, ANNs suffer from difficulty in training predictors and may not, therefore, produce a unique solution over various runs due to different weights [COULIBALY, EVORA 2007; KHAN, COULIBALY 2006].

As a result, the present study sought to explore, for the first time, the use of ELM, ELMw and ELMwb water demand forecasting models for short-term urban water demand forecasting for the city of Calgary (Alberta Canada), and compare their performance to that of previously applied ANN, ANNw, and ANNwb models [TIWARI, ADAMOWSKI 2015a, b]. As only three years of urban water demand data were available for calibration and validation of the models, a secondary aim of this study was also to explore how these methods fared in situations with limited data. The input variables applied in this study consisted of average daily water demand, maximum temperature and total precipitation.

THEORETICAL OVERVIEW

EXTREME LEARNING MACHINE

Owing to its prior application in hydrology [ACHARYA et al. 2014; DEO, ŞAHIN 2015a], the present study has extended the application of ELM algorithm-based models [HUANG et al. 2006] to forecasting daily urban water demand (UWD). Based on state-of-the-art single-layer feed-forward network algorithms, ELMs are similar to feed-forward back-propagation ANNs (ANNF) and least square support vector regression (LSSVR). However, compared to the latter algorithms, ELMs have greater ability to
solve regression problems efficiently in a short modelling time [HUANG et al. 2012] and show a relatively better predictive performance [ACHARYA et al. 2014; DEO, ŞAHIN 2015a, b]. Moreover, in ELMs the weights and hidden neuron selections are randomized, so that the output weights have a unique least-square solution solved by way of the Moore-Penrose generalized inverse matrix method [HUANG et al. 2006]. Consequently, the ELM is a simple, three-step procedure requiring no parameterization except the randomized determination of hidden neurons. Optimisation is performed by choosing activation functions for hidden nodes based on sigmoid, radial basis or hard limit equations [DEO, ŞAHIN 2015a, b; ŞAHIN 2012; ŞAHIN et al. 2013; 2014]. This yields distinct advantages over conventional models.

Figure 1(a) illustrates the general ELM modelling framework. Mathematically, for a set of predictive samples \((X_i, t_i)\) where \(i = 1, 2, \ldots, N\) with inputs, \(X_i = [x_{i1}, x_{i2}, \ldots, x_{im}] \in \mathbb{R}^m\) and \(t_i \in \mathbb{R}\), the ELM architecture consists of \(L\) random hidden neurons with activation function \(g(x)\) such that [HUANG et al. 2006]:

\[
\sum_{z=1}^{L} \beta_z g(z) (X_i) = \sum_{z=1}^{L} \beta_z g(a_z X_i + c_z) = O_i \quad (1)
\]

**Fig. 1.** An illustration of (a) extreme learning machine (ELM) and (b) artificial neural network (ANN) modelling frameworks used for prediction of the UWD; source: own study

Where \(a_z = [a_{z1}, a_{z2}, \ldots, a_{zm}]^T\) is the weight vector connecting the \(z^{th}\) hidden neuron, the input neuron, \(\beta_z = [\beta_z]^T\) is the weight vector connecting the \(z^{th}\) hidden neuron and the output neuron, \(O_i\) (which is the same dimension as the target function, \(T_i\)), \(c_z\) is the threshold of the \(z^{th}\) hidden neuron, and \(a_z X_i\) denotes the inner product. The output neurons, which are considered to be linear in this study, do not require any transformative equation. It has been proven by HUANG et al. [2006] that Single-hidden-layer feed-forward ANNs with \(L\) hidden nodes and an activation function \(g(x)\) have the proven capacity to approximate \(N\) training pair samples with zero error,

Thus there must exist \(\beta_z, a_z\) and \(b_z\) such that:

**Fig. 1.** An illustration of (a) extreme learning machine (ELM) and (b) artificial neural network (ANN) modelling frameworks used for prediction of the UWD; source: own study

Where \(a_z = [a_{z1}, a_{z2}, \ldots, a_{zm}]^T\) is the weight vector connecting the \(z^{th}\) hidden neuron, the input neuron, \(\beta_z = [\beta_z]^T\) is the weight vector connecting the \(z^{th}\) hidden neuron and the output neuron, \(O_i\) (which is the same dimension as the target function, \(T_i\)), \(c_z\) is the threshold of the \(z^{th}\) hidden neuron, and \(a_z X_i\) denotes the inner product. The output neurons, which are considered to be linear in this study, do not require any transformative equation. It has been proven by HUANG et al. [2006] that Single-hidden-layer feed-forward ANNs with \(L\) hidden nodes and an activation function \(g(x)\) have the proven capacity to approximate \(N\) training pair samples with zero error,

Thus there must exist \(\beta_z, a_z\) and \(b_z\) such that:
\[
\sum_{i=1}^{c-L} \beta_i g(a_i X_i + c_i) = T_i
\]  \tag{2}

Which can be written in a compact form as

\[
H\beta = T,
\]

where

\[
H(a_1, \ldots, a_L, c_1, \ldots, c_L, X_1, \ldots, X_L) = \begin{bmatrix}
(a_1 X_1 + c_1) & \cdots & (a_1 X_N + c_1) \\
\vdots & \ddots & \vdots \\
(a_L X_1 + c_L) & \cdots & (a_L X_N + c_L)
\end{bmatrix}_{N \times L}
\]  \tag{3}

and

\[
\beta = \begin{bmatrix}
\beta_1^T \\
\vdots \\
\beta_L^T
\end{bmatrix}_{L \times 1} \quad \text{and} \quad T = \begin{bmatrix}
T_1^T \\
\vdots \\
T_N^T
\end{bmatrix}_{N \times 1}
\]  \tag{4}

It is noteworthy that the input weights and hidden neuron biases are randomly generated in the ELM model, which is different from the ANN that requires iterative tuning of parameters, and thus, requires greater modelling time [DEO, SAHIN 2015a; ŞAHIN et al. 2013; 2014]. The training algorithm is used to find least squares solutions to the system of equations \(H\beta = T\) and the parameter \(\beta\) can directly be determined as

\[
\bar{\beta} = H^+ T
\]  \tag{5}

where the \(\bar{\beta}\) is the smallest least-square linear system solved using the \(H^+\) as the Moore-Penrose generalized inverse of \(H\).

In order to develop an ELM model using a set of predictive samples \((X_i, t_i)\), the forecasts of the UWD, \(UWD_h(t)\) are given by:

\[
UWD_h(t) = \sum_{i=1}^{L} \beta_i g(a_i X_i + c_i)
\]  \tag{6}

In the present study, the time-series forecasts denoted by \(UWD_h(t)\) were generated using activation functions \(g(X)\) described by the logarithmic sigmoid, \(\psi(X)\) and the output function, \(\chi(X)\) equations [VOGL. et al. 1988] as seen in equations (7) and (8).

\[
\psi(X) = \frac{1}{1 + e^{-X}}
\]  \tag{7}

\[
\chi(X) = \text{linear}(X)
\]  \tag{8}

**ARTIFICIAL NEURAL NETWORK**

A well-established class of nonlinear modelling techniques mimicking the biological functions of the human brain [McCulloch, Pitts 1943], ANN models served as a benchmark for ELM model performance in predicting UWD in this study. Basically, an ANN represents a highly interconnected framework that sends information from an input to an output layer through weighted connections and functional neurons to facilitate nonlinear mapping of the predictive dataset to high-dimensional hyper-planes, as demonstrated in Fig. 1(b). This allows the separation of data patterns, formation of idealised models and subsequent UWD predictions.

Widely applied in hydrology, the popular ANN\_type class of ANN models, equipped with multi-layer perceptron functional neurons [ABBOT, MAROHASY 2012; ADAMOWSKI et al. 2012c; DEO, ŞAHIN 2015a; KESKIN, TERZI 2006; MEKANIK et al. 2013] was used in the present study. The ANN architecture is designed to successively update the model parameters (weighted connections and neuron biases) to drive the empirical error to a set tolerance through each iteration (epochs) of forward passing of updated parameters and backward propagation of the errors to tune them.

For a set of predictive (input) sample \((X_i, t_i)\) where \(i = 1, 2, \ldots, N\) denotes the sequence of inputs, \(X_i = [x_{i1}, x_{i2}, \ldots, x_{in}] \in \mathbb{R}^n\) and \(t_i \in \mathbb{R}\), the FFBP-ANN model formulated is written as [KIM, VALDES 2003]:

\[
UWD_h(t) = f_o \left[ \sum_{j=1}^{L} w_{jo} \right] \left[ \sum_{i=1}^{N} w_{ji} X_i(t) + w_{jo} + w_{io} \right]
\]  \tag{9}

where, \(L\) (determined iteratively rather than randomly as with the ELM model) is the number of hidden neurons, \(X_i(t)\) is the \(i\)th input variable at the time-step, \(w_{ji}\) is the weight that connects the \(j\)th neuron in input layer and the \(i\)th neuron in the hidden layer and \(w_{jo}\) is the bias for the hidden \(j\)th hidden neuron.

In literature, second-order training methods with Levenberg-Marquardt (LM) and Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton backpropagation algorithms are used [DENNIS, SCHNABEL 1996; MARQUARDT 1963]. The algorithm is used to minimize the mean squared error of the predicted and observed UWD [TIWARI, ADAMOWSKI 2013]. In our study, an LM algorithm that uses an approximation to the Hessian matrix was used as follows [DEO, ŞAHIN 2015a]:

\[
x_{k+1} = x_k - [J^T J + \mu I]^{-1} J^T e
\]  \tag{10}

where \(J\) is the Jacobian matrix calculated using standard backpropagation techniques and is less complex than computing the Hessian matrix [MARQUARDT 1963]. The \(J\) contains first derivatives of network errors with respect to the weights and biases, \(e\) is a vector of errors, \(\mu\) is the combination coefficient and \(I\) is the identity matrix.

**DISCRETE WAVELET TRANSFORMATION**

A primary purpose of this study was to demonstrate the effectiveness of ELM\_W models for urban water demand forecasting. In general, wavelet de-
composition is a multi-resolution tool for pre-processing of non-stationary signals. This is similar to short-time Fourier transformation as a windowing technique in which the time-series are decomposed into the shifted and scaled versions of a wavelet, termed the mother wavelet. These can serve in extracting frequency-based information from current time-series that can then be used to predict future time-series. Assuming a continuous time-series \( X(t) \) as an input vector where \( t \in [\infty, -\infty] \), a wavelet function \( \psi(\eta) \) that depends on a non-dimensional time parameter \( \eta \) is defined as:

\[
\psi(\eta) = \Psi(\tau, s) = s^{-\frac{1}{2}} \psi \left( \frac{t - \tau}{s} \right)
\]

(11)

where \( t = \) time, \( \tau = \) time step in which the window function is iterated and \( s \in [0, \infty] \) is the wavelet scale. The term \( \psi(\eta) \) must have zero mean and localized in both the time and Fourier space.

The discrete wavelet transformation (DWT) selects translation and location parameters for the input signal. Subsequently, discrete wavelets coefficients (DWCs) are obtained that represent the minimum number of components needed to reflect the time-series according to the mother wavelet. Several families of wavelets have proven useful in a range of applications [MALLAT 1989]. For practical applications, hydrologists use wavelets to analyse a discrete rather than a continuous signal. This discrete wavelet is of the form:

\[
g_{i,j}(t) = \frac{1}{\sqrt{a_0}} \left( \frac{t - jb_0d_0^i}{a_0} \right)
\]

(12)

where \( i \) and \( j \) are the integer values, and \( b_0 \) and \( a_0 \) are the location parameter and the specified fine dilation step, respectively. Common values for \( a_0 \) and \( b_0 \) are 2 and 1, respectively [SEHGAL et al. 2014; TIWARI, ADAMOWSKI 2013]. The discrete wavelet transform involves selecting scales and positions based on powers of two, (called the dyadic scales and translations). The dyadic wavelet can be compressed as follows:

\[
g_{i,j} = 2^{-j/2} g(2^{-i} h - f)
\]

(13)

\[
T_{i,j} = 2^{-j/2} \sum_{h=0}^{2^j-1} g(2^{-i} h - f)x_h
\]

(14)

where \( T_{i,j} \) is the wavelet coefficient for the discrete wavelet of scale \( a=2^i \) and the location \( b = 2^j \). Equation (14) considers a finite time series, \( x_h, h = 0, 1, 2, \ldots, j-1 \); and \( j \) is an integer power of 2, i.e., \( j = 2^j \).

The inverse discrete transform is given by:

\[
x_h = \sum_{i=1}^{j} \sum_{j=0}^{2^j-1} T_{i,j} 2^{-i/2} g(2^{-i} h - f) = T + \sum_{i=1}^{j} W_i(t)
\]

(15)

where \( T \) is called the approximation sub-signal at level \( i \), and \( W_i(t) \) represents the details of the sub-signals at level \( i = 1, 2, \ldots, l \).

The wavelet coefficients, \( W_i(t) (i = 1, 2, \ldots, l) \) provide the details of the signal, which can capture small features of interpretational values in the inputted dataset. The residual term, \( T \), represents the background information of the data. The wavelet is robust since it does not include any potentially erroneous assumption or parametric testing procedure. Because of the simplicity of \( W_i(t) \), the relevant characteristics in the hydrologic dataset (e.g. periods, hidden periods, dependence and jumps) can be diagnosed through these discrete wavelet components [TIWARI, CHATTERJEE 2011]. Consequently, the prediction accuracy of drought models are improved.

**BOOTSTRAP TECHNIQUE**

There are three sources of uncertainty that affect the output of the ANN and ANNw models: parameter uncertainty, sub-optimal training and insufficient input variables. Bootstrapping is a computational, data-driven simulation method that can be used to assess uncertainty by measuring the variance \( \sigma^2 \) of \( S \), the bootstrap resample. Bootstrap samples are generated through an intensive resampling with replacement method. These samples or realizations provide a better understanding of the mean and variability of the original data, and thus of its unknown distribution or process, thereby reducing uncertainty [EFRON 1979; EFRON, TIBSHIRANI 1993].

Assume a population \( T \) with unknown probability distribution \( F \), where \( t=(x,y) \) is a realization drawn independently and identically distributed (i.i.d.) from \( T \), \( x \) is an input vector and \( y \) is the corresponding output vector, and \( n \) is the size of original dataset. In this case:

\[
T_n = \{(x_1,y_1), (x_2,y_2), \ldots, (x_n,y_n)\}
\]

(16)

is a bootstrap resample obtained from an empirical distribution function, \( F \) with a mass of \( 1/n \) for each \( t_1, t_2, \ldots, t_n \). Similarly, a set of bootstrap samples such as \( T^1, T^2, \ldots, T^n \) can be produced, where \( s \) is a particular bootstrap resample, whereas \( S \) is the total number of bootstrap resamples. In such a case the total number of bootstrap samples, \( S \), usually ranges from between 50 and 200 samples [EFRON 1979].

In this study, several bootstrap resamples were generated and used to train several different ANN and ANNw models. Ensemble forecasts were obtained and designated ANNw and ANNwB, respectively. For each \( T \), an ANN and ANNw model was developed and trained using all \( n \) observations. The output of ANN models was represented in terms of bootstrap resamples and corresponding optimized weights as \( f_{\text{opt}}(x_0, w) \) where \( x_0 \) was the input data pattern, and \( w \) was the optimized weights of the ANN model for a particular bootstrap resample \( s \).
both the models was then evaluated using a set \( A_0 \). Then the generalization error denoted as \( \hat{E}_0 \) was estimated (e.g., ANN model) as [TWOMEY, SMITH 1998]:

\[
\hat{E}_0 = \frac{\sum_{s=1}^{S} \sum_{t \in A_s} \left[ y_t - f_{NN}(x_t, w_s) \right]^2}{\sum_{s=1}^{S} \#(A_s)} \tag{17}
\]

\( A_s \) is a set of observation pairs \( t = (x_t, y_t) \) that were not included in generating the bootstrap resamples or it is the set of data patterns in the testing data set or the dataset not included in the development of different resamples. \( S \) is the number of bootstrap samples generated from the training dataset.

Finally, the estimate \( \hat{y}(x) \) of the ANN\(_B\), ANN\(_W\) and ANNB were presented as the average of the \( S \) bootstrapped estimates of the corresponding ANN model (e.g., ANN\(_B\)) as [TIWARI, CHATTERJEE 2011]:

\[
\hat{y}(x) = \frac{\sum_{s=1}^{S} f_{NN}(x_s, w_s)}{S} \tag{18}
\]

and the variance using \( S \) resample was given by:

\[
\sigma^2(x) = \frac{\sum_{s=1}^{S} \left[ f_{NN}(x_s, w_s) - \hat{y}(x) \right]^2}{S - 1} \tag{19}
\]

A number of forecasts obtained with ANN and ANN\(_W\) models, trained with multiple realizations of the training dataset, served to generate a 95% confidence interval (CI), i.e., two tailed \( \alpha = 0.05 \) significance level. These CIs indicated the frequency with which the CIs would contain the true value in the repeated application of the model. A 100\( \cdot (1 - \alpha) \)% percent CI covering the overall UWD water demand \( \hat{y}(x) \) can be estimated as [EFRON, TIBSHIRANI 1993]:

\[
CI = (LB, UB) = \left[ \hat{y}(x) + t_{\alpha/2}^2 \sigma(x), \hat{y}(x) + t_{\alpha/2}^2 \sigma(x) \right] \tag{20}
\]

where \( n \) is the total number of water demand observations, \( p \) is the total number of parameters in the NN and WNN models, \( t_{\alpha/2}^2 \) is the \( \alpha/2 \) percentile for the Student distribution, with \( n - p \) degrees of freedom, UB is the upper bound, LB is the lower bound, and \( \sigma(x) \) is the standard deviation of \( S \) bootstrapped estimates.

**PERFORMANCE INDICES**

The developed models’ performances were evaluated using five statistical indices: the coefficient of determination \( (R^2) \), root mean square error \( (RMSE) \), persistence index \( (PI) \), mean absolute error \( (MAE) \), and peak percentage deviation \( (P_{dv}) \), as defined below [DAWSON et al. 2007].

(i) The coefficient of determination \( (R^2) \) performance index is a squared ratio of the combined dispersion of two time series to the total dispersion of the observed and modelled time series. It presents the overall agreement between observed and modelled time series and varies from 0 for a poor model to 1 for a perfect model.

The coefficient of determination \( (R^2) \) is expressed as:

\[
R^2 = \frac{\sum_{i=1}^{n} (O_i - \bar{O})(P_i - \bar{P})^2}{\sqrt{\left( \sum_{i=1}^{n} (O_i - \bar{O})^2 \right) \left( \sum_{i=1}^{n} (P_i - \bar{P})^2 \right)}} \tag{21}
\]

where \( n \) is the number of data points, \( O_i \) and \( P_i \) are the \( i^{th} \) observed and \( i^{th} \) forecasted UWD values, respectively, and \( O \) and \( P \) are the observed and forecasted UWD means, respectively.

(ii) The root mean square error \( (RMSE) \) is a measure of overall performance across the entire range of the dataset and provides a good measure of model performance for high flows [KARUNANITHI et al. 1994], as it is sensitive to small differences in model performance and exhibits high sensitivity to the larger errors occurring for higher magnitudes. It is expressed as:

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (O_i - P_i)^2} \tag{22}
\]

The \( RMSE \geq 0 \), and shows a perfect model fit \( RMSE = 0 \).

(iii) Percentage deviation in peak \( (P_{dv}) \). It is defined as:

\[
P_{dv} = 100 \left( \frac{P_{O\text{peak}} - O_{\text{peak}}}{O_{\text{peak}}} \right) \tag{23}
\]

where, \( O_{\text{peak}} \) and \( P_{\text{peak}} \) are the peak of observed and forecasted water demand, respectively.

(iv) The mean absolute error \( (MAE) \) measures overall agreement between observed and forecasted values, but is not weighted towards higher or lower magnitude events. It evaluates all deviations from the observed values equally, without considering sign. So \( MAE \geq 0 \), with \( MAE = 0 \) representing a perfect model fit to observed values. It is expressed as:

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |O_i - P_i| \tag{24}
\]
(v) The persistence index (PI) is one minus the ratio of the sum square error (SSE) to the same SSE obtained when the last observed value itself is considered as the forecasted value for a particular lead time. The more PI exceeds zero the greater the model’s accuracy; however, if PI = 0 then the model has performed no better than a one parameter 'no knowledge' model, while if PI < 0 the model has performed more poorly than a 'no knowledge' model. It is essentially a comparison between the model under study and a simple naïve persistence model. Thus, to estimate PI the predicted water demand at time \( t (P_t) \) is considered as the observed water demand at time \( t - j \) \( (O_{t-j}) \), where \( j \) indicates the lead time selected for the water demand forecast. PI is therefore expressed as:

\[
PI = 1 - \frac{\sum (O_t - P_t)^2}{\sum (O_t - O_{t-j})^2} = 1 - \frac{SSE}{SSE_{naive}} \tag{25}
\]

MATERIALS AND METHODOLOGY

STUDY AREAS AND DATA PARTITIONING

With a population of approximately 1.1 million people, Calgary is amongst the largest cities in Canada [City of Calgary 2011]. The Bearspaw Plant treats water from the Bow River primarily to supply the northern half of the city, while the Glenmore Plant treats water from the Elbow River and supplies the southern portion of the city [City of Calgary 2011]. Each plant supplies about half of Calgary’s total drinking water needs, and the 4600 km distribution system is interconnected through transmission mains. Since 1980, the city has invested in maintenance of the network by replacing corroded pipes with PVC and by adding cathodic protection on pipes to reduce the rate of corrosion. As a result, emergency repairs have been reduced by 73% [City of Calgary 2011].

In 2010, the total per capita water demand in Calgary was 406 L·d\(^{-1}\), while residential use was 257 L·d\(^{-1}\), less than the Canadian average of 343 L·d\(^{-1}\) but still greater than other Prairie cities [Environmental and Safety Management 2010]. The Government of Alberta announced in 2006 that new water licenses for the Bow River Basin would no longer be granted, which has led to an increased awareness and need for water conservation and a demand for reduction measures in the midst of continuing population growth and climate change. For example, the Calgary City Council has adopted a goal of reducing total per capita use to 350 L·d\(^{-1}\) by 2033, metering all residential homes by the end of 2014, and maintaining peak demand below 0.95·10\(^{6}\) L through to 2032 [Environmental and Safety Management 2010].

The average summer high in Calgary is 20°C, with a historic extreme high of 36°C, and the average winter low is –13°C, with a historic extreme low of –45°C. Annual rainfall in Calgary is about 320 mm, with a recorded extreme daily rainfall of 95.3 mm. Annual snowfall averages around 125 cm, with an extreme daily snowfall of 48.4 cm [Environment Canada 2010].

The data obtained from the City of Calgary consisted of average daily water demand, maximum temperature, and total precipitation, compiled from 25.03.2004 to 31.12.2006. Additional data were not available. For the development of the models, the data was divided into three sets: one for training the models, one for cross-validation to check that the models did not over-fit, and one for testing the performance of the developed models. The details of the data partitioning are shown in Table 1.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Period</th>
<th>Number of data patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>25.03.2004 to 24.03.2005</td>
<td>365</td>
</tr>
<tr>
<td>Cross-validation</td>
<td>25.03.2005 to 24.03.2006</td>
<td>365</td>
</tr>
<tr>
<td>Testing</td>
<td>25.03.2006 to 31.12.2006</td>
<td>282</td>
</tr>
</tbody>
</table>

Source: own study.

INPUT SELECTION AND DROUGHT MODEL DEVELOPMENT

ANN Model Development

Selection of significant input variables and identifying optimal model structure are two important steps in ANN model development. Correlation statistics (e.g., cross-correlation, auto-correlation and partial auto-correlation) along with a trial and error approach were employed to obtain prior knowledge of input variables dynamics. In this procedure information at different lag times of daily UWD, daily maximum temperature and daily total precipitation were considered. Following this, the optimal network geometry for the ANN model was identified by trial and error, and the number of hidden neurons that produced the lowest generalization error, ranging between 1 and 15, was considered to be the optimal structure [Jia, Culver 2006]. ANN models were initially developed using the significant inputs that were log-transformed and linearly scaled to a range of 0 to 1 [Campolo et al. 1999]. A second-order training method, the Levenberg–Marquardt optimization method was used to minimize the mean square error (MSE) between the forecasted and observed UWD values.

ELM, ELM\(_b\) and ELM\(_w\) Model Development

Based on an earlier study that demonstrated the practical use of ELM models for drought forecasting in eastern Australia [Deo, Şahin 2015a, b], a 3-layer network containing input, feature optimisation and output spaces was employed (Fig. 1a, b). The ELM model employed in the current study was developed using the logarithmic sigmoid activation function.
Initially, the ELM model was randomly executed ~50–1000 times to explore input layer weights, weights and optimal nodes in the hidden layer and model biases yielding the smallest MSE. This resulted in ~100 randomisations yielding a stable solution. In each case, the CPU time consumed to run urban water demand models was recorded.

Also shown to be an effective tool in drought forecasting [DEO, SÄHIN 2015b], an ANN model was developed as a benchmark. The ANN model was stopped early when processing the validation dataset to avoid overtraining or over-fitting [ADAMOWSKI 2008a, b; ADAMOWSKI et al. 2012a; TIWARI, ADAMOWSKI 2013]. During this process the MSE was monitored at iterations of the training and during cross-validation phases. The training was stopped when the MSE reached a minimum [BISHOP 1995]. As with previous studies [DEO, SÄHIN 2015a, b; TIVARI, ADAMOWSKI 2013], the fast and efficient second-order Levenberg–Marquardt training algorithm was employed in the ANN model.

A robust ELM model was developed by considering different input variables and optimization parameters. Two further hybrid models were developed: bootstrap-based ELM (ELM B), and wavelet-based ELM (ELMW). MATLAB® (v.7.10.0) code was written to develop all the wavelet models, while bootstrap resamples were generated using an Excel add-in (Bootstrap.xla) [BARRETO, HOWLAND 2006]. The ensemble of roughly 200 ELM_B models were developed for each bootstrap resample dataset, and all the 200 forecasts were later combined to generate an ensemble of all these forecasts. To further improve the ELM, we applied discrete wavelet transformation (DWT) on the predictor signals to achieve a timescale representation of the localized and transient phenomenon at different scales in the data series [ADAMOWSKI 2008a, b; ADAMOWSKI, CHAN 2011; ADAMOWSKI et al. 2012a, b; KIM, VALDES 2003; TIVARI, ADAMOWSKI 2013]. The DWT process aimed to achieve a time-scale realisation of both the localized and the transient phenomena at various frequencies. The frequency content and temporal variations was analysed by effectively decomposing inputs into discrete wavelet coefficients (DWCs) to make the non-stationarity obvious and the model more responsive to the variations in frequencies of the input data [TIVARI, ADAMOWSKI 2013].

The utilized wavelet function was adopted from a family of the Daubechies mother wavelet [NOURANI et al. 2009; WU et al. 2009] whereby the DWT process operated as two sets of functions with a high-pass and a low-pass filter. The predictor variables were passed through the high- and low-pass filters to acquire detail (db1, db2, db3) in terms of high frequency components and approximation coefficients (A3) in terms of low frequency components of the signal. As the performance of the db5 wavelet with three levels of decomposition provided the best performance, for illustrative purposes only 3 levels of decomposition (db1, db2, db3) and 1 approximation (A3) are presented for the UWD data over the tested period (Fig. 2). The low-frequency components reflected by A3 showed the broad-scale patterns in the predictor dataset including its periodicity and trends, and was closely in-phase with the predictor signal, whereas the high-frequency components (db1, db2, db3) appeared to replicate greater details of the subtle but significant patterns in the UWD time-series [KÜÇÜK, AğIRALI‘OĞLU 2006].

Though earlier studies have demonstrated a better performance of wavelet-based models, the ways in which the wavelet sub time-series are included in model development can vary greatly. Some studies have used all of their wavelet sub-series [ADAMOWSKI, SUN 2010; NOURANI et al. 2009; WANG, DING 2003] whereas others have removed the db1 sub-series and added the remaining series, considering the former series as noise due to its low correlation with their original data [KİŞİ, ÇİMEN 2011; PARTAL, KİŞİ 2007; RAJAEE et al. 2010]. However, in some studies, new wavelet time-series were developed by adding up the effective DWCs based on regression correlation [TIVARI, CHATTERJEE 2010b; 2011]. As it is wise not to completely rely on a model based on a particular wavelet series that captures some phenomena at the expense of others [RATHINASAMY et al. 2013], we considered each wavelet function in terms of its own strengths in capturing stochastic characteristics and physical structure of the hydrological dataset.

**ANN_B and ANN_W Model Development**

ANN_B models were developed in similar manner as ELM_B models. The ANN_W models were developed by inputting the wavelet sub time series produced using DWCs. In this study, wavelet functions from the Daubechies family of wavelets [NOURANI et al. 2009; WU et al. 2009] were used, and three levels of decomposition were considered based on the following formula [NOURANI et al. 2008]:

\[ L = \text{int} \left\{ \log(n) \right\} \]  

(26)

where, \(L\) is the number of decomposition levels, and \(n\) is the number of time series data.

The number of datasets \(n = 1012\) yield a value of \(L = 3\), leading to three levels of decomposition (d1, d2, and d3) and approximation (A3) for the data (Fig. 2). The effective DWCs were determined using the correlation coefficients between each wavelet component and the observed UWD. The correlation between the original daily time series for Calgary and corresponding different wavelet sub-time series are shown in Table 2.

In earlier studies [ADAMOWSKI, SUN 2010; KİŞİ 2010; TIVARI, CHATTERJEE 2010a; 2011], the significant wavelet sub-time series of a particular time series was used and added to generate a new time series,
Fig. 2. Wavelet sub-time series of the (a) daily water demand and (b) daily maximum temperature of Calgary from 24 March, 2004 to 31 December, 2006; A3 = approximation, d1, d2, d3 = details; source: own study
RESULTS AND GENERAL DISCUSSION

WATER DEMAND FORECASTING IN CALGARY USING ELM, B-ELM AND W-ELM MODELS

For 1 day lead-time UWD forecasting, the significant input variables were water demand at time $t$ [i.e., WatDemand($t$)] and maximum temperature at time $t$ [i.e., MaxT($t$)]. The statistical and graphical (scatter plots) assessment of ELM models’ performance in UWD forecasting are presented in Table 3 and Figure 3a, respectively. Statistical performance metrics for the ELM model were generally satisfactory, indicating that the margin of difference between actual and forecasted UWD was relatively small. Considering the range of water demand in Calgary (313.38 mL·d$^{-1}$ ≤ UWD ≤ 684.25 mL·d$^{-1}$), the performance of the best ELM models ($RMSE = 33.02$ mL·d$^{-1}$), can be viewed as satisfactory for 1-day lead-time UWD forecast.

Table 3. Water demand forecasting for 1 day lead times using the best models for testing dataset

<table>
<thead>
<tr>
<th>Model</th>
<th>Lead time</th>
<th>$R^2$</th>
<th>RMSE mL·day$^{-1}$</th>
<th>$P_a$ %</th>
<th>MAE mL·day$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELM</td>
<td>1</td>
<td>0.850</td>
<td>33.02</td>
<td>10.06</td>
<td>24.42</td>
</tr>
<tr>
<td>B-ELM</td>
<td>1</td>
<td>0.851</td>
<td>32.97</td>
<td>9.61</td>
<td>24.14</td>
</tr>
<tr>
<td>W-ELM</td>
<td>1</td>
<td>0.927</td>
<td>23.11</td>
<td>6.04</td>
<td>16.70</td>
</tr>
</tbody>
</table>

Source: own study.

This study further explored the capacity of bootstrap-based ensemble modelling along with ELM models for UWD forecasting. Some 200 bootstrap resamples of the training dataset were generated and ELM$_W$ models were developed by averaging the forecasts obtained from 200 resultant ELM. The statistical and graphical (scatter plots) assessment of ELM$_W$ models’ performance in UWD forecasting are presented in Table 3 and Figure 3b, respectively. On the basis of these assessments the ELM$_W$ could be seen to slightly outperform the ELM model. Given that the ELM$_W$ model was developed using several realization of the training dataset it was expected to produce stable and reliable results even if the pattern of the training and testing datasets changed.

The efficacy in daily UWD forecasting of ELM$_W$ models relative to other machine learning models was also considered. In order to develop the ELM$_W$ models, all the time series datasets were decomposed into approximation and details and all the components were considered separately as inputs for ELM$_W$ model development. The ELM$_W$ models were developed using wavelet sub-time series derived from the dataset that produced the best ELM model for Calgary. Based on a trial and error process the best ELM$_W$ models for 1-day lead-time forecasts were found with inputs $a_3(t)$, $d_3(t)$, $d_2(t)$, and $d_1(t)$ at time $t$ of wavelet components $A_3$, $d_1$, $d_2$, $d_3$ of daily water demand (WatDemand), daily maximum temperature (MaxT) and total precipitation (TotP), respectively. The $d_1$ component showed a lesser correlation with the original water demand than the other components (i.e., $A_3$, $d_2$, and $d_3$), indicating that $d_1$ components may be redundant/noise information contained in the original water demand time series (Tab. 2). However, it was found to not be altogether without importance during UWD forecasting, as it showed that all the components exhibit some important information about the physical characteristics of the original UWD time series. The statistical and graphical (scatter plots) assessment of ELM$_W$ models’ performance in UWD forecasting are presented in Table 3 and Figure 3c, respectively. The scatter plots revealed the ELM$_W$ model to significantly outperform the unenhanced ELM model ($R^2 = 0.927$ vs. $R^2 = 0.850$, respectively) in UWD forecasting. This showed the supremacy of ELM$_W$ models in UWD forecasting compared to ELM and ELM$_B$ models.

WATER DEMAND FORECASTING IN CALGARY USING ANN, B-ANN AND W-ANN MODELS

In an earlier study Tiwari and Adamowski [2015] developed models for UWD forecasting by applying the same input variables for the same dataset length. One can therefore compare the performance of newly developed ELM, ELM$_M$ and ELM$_W$ models with the earlier applied ANN, ANN$_N$ and ANN$_W$ models. The structure and performance of the best ANN models for the UWD forecasting testing dataset for Calgary are shown in Table 4. For 1-day lead-time UWD forecasting, the significant input variables obtained were water demand at time $t$ [i.e., WatDemand($t$)] and maximum temperature at time $t$ [i.e., MaxT($t$)]. The number of optimum hidden neurons was identified as 3.

The hydrographs and scatter plots of observed and forecasted UWD values using the best ANN models for 1 day lead-time UWD prediction (Fig. 4a) show that ANN models’ performance to be acceptable, with forecasted values following the general trend.
Water demand forecasting using extreme learning machines

Fig. 3. Hydrographs and scatter plots for observed (Obs) and predicted (Pred) water demand in Calgary for 1 day lead time forecasts for the testing dataset using the best models: a) ELM, b) B-ELM, c) W-ELM; source: own study

Table 4. Water demand forecasting for 1 day lead times using the best models for testing dataset

<table>
<thead>
<tr>
<th>Model</th>
<th>Lead time</th>
<th>$R^2$</th>
<th>RMSE mL·day$^{-1}$</th>
<th>$P_e$ %</th>
<th>MAE mL·day$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>1</td>
<td>0.860</td>
<td>32.97</td>
<td>11.92</td>
<td>24.31</td>
</tr>
<tr>
<td>B-ANN</td>
<td>1</td>
<td>0.854</td>
<td>33.71</td>
<td>13.47</td>
<td>24.67</td>
</tr>
<tr>
<td>W-ANN</td>
<td>1</td>
<td>0.924</td>
<td>24.15</td>
<td>8.25</td>
<td>17.48</td>
</tr>
</tbody>
</table>

Source: own study.

of observed values and yielding an almost 1:1 regression line in forecasted vs observed scatter plots. The performance of ANN models was comparable to that of ELM models; however, the ELM models were more time efficient than the ANN models (ELM ≈ 2 sec, ELM$_B$ ≈ 400 sec, ELM$_W$ ≈ 1 sec, ANN ≈ 4 sec, ANN$_B$ ≈ 800 sec and ANN$_W$ ≈ 3 sec CPU time).

Similar to ELM$_B$ models, the ANN$_B$ models were developed using bootstrap resamples of the training dataset used to develop the best ANN models. For each lead time, results from 200 ANN models developed from 200 bootstrap resamples were averaged to generate the ANN$_B$ forecast. Based on observed vs. forecasted UWD scatter plots (Fig. 4b), and statistical performance indices (Tab. 4) for the testing dataset the performance of the ANN$_B$ models to be comparable to that of ANN models for 1-day lead-time UWD forecasting, but their accuracy decreases significantly for longer lead-times. The performance of the ANN$_B$ models was slightly less accurate than those of the ANN models. Likewise, the performance of ELM$_B$ model was slightly better than that of the ANN$_B$ model for UWD forecasting.

The statistical and graphical (scatter plots) assessment of ANN$_W$ models’ performance in UWD forecasting with 1-day lead-times are presented in Table 4 and Figure 4c, respectively. The best ANN$_W$ model outperformed the best ANN and ANN$_B$ models, demonstrating the ability of wavelet analysis to capture useful information from different periodic components (i.e. wavelet sub-time series). The ELM$_W$ model performed slightly better than the ANN$_W$ models for UWD forecasting.
CONCLUSIONS

Accurate and reliable UWD forecasting is necessary to help transition to more effective and sustainable urban water resources planning and management [Butler, Adamowski 2015; Halbe et al. 2013; 2014; Inam et al. 2015; Kolinjivadi et al. 2014; Straith et al. 2014]. In this study, ELM\textsubscript{W} models based on their capacities of wavelet transformation and ELM modeling techniques were employed to simulate the UWD in the city of Calgary, Canada. A limited yet more appropriate set of predictor variables were utilized. The feasibility of using ELM\textsubscript{W} for UWD forecasting was compared to that of traditional ELM and ANN models, as well as ANN\textsubscript{B} and ANN\textsubscript{W} models. In this study ELM, ANN, ELM\textsubscript{B}, ANN\textsubscript{W} and ANN\textsubscript{B} models were developed for 1-day lead time UWD forecasting for the city of Calgary, Canada. Based on five performance indices ($R^2$, $P_{dv}$, RMSE, $PI$, MAE), ELM\textsubscript{W} models were found to perform considerably better than ANN, ELM, ANN\textsubscript{B}, and ANN\textsubscript{W} models. This highlights the ability of wavelet transformation to decompose time-series data with non-stationarity into discrete wavelet components, highlighting cyclic patterns and trends in the time series data at varying temporal and spatial scales and making the data readily usable in forecasting. Indeed, the margins of prediction errors were much smaller for the ELM\textsubscript{W} and the model execution time was shorter compared to the other machine learning models considered in this work. Therefore, as a pioneer study on the application of ELM\textsubscript{W} modelling to UWD prediction, this research clearly demonstrated the feasibility of wavelet-based modelling for UWD forecasting. Moreover, this study provides a promising advancement to machine learning models for UWD studies and the opportunity to explore the ELM and wavelet techniques in real-time UWD forecasting.

REFERENCES


DAWSON C.W., ABRAHART R.J., SEE L.M. 2007. HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. Envi-


Khan M.S., Coulthay P. 2006. Application of support vector machine in lake water level prediction. Journal of
Przewidywanie zapotrzebowania na wodę z użyciem technik uczenia maszynowego

STRESZCZENIE

Słowa kluczowe: bootstrap, ekstremalne maszyny uczące się, falki, Kanada, niepewność, prognozowanie zapotrzebowania na wodę, sztuczne sieci neuronowe

Oceniono zdolność modelowania z użyciem ekstremalnej maszyny uczącej się (ELM) stosowanej samodzielnie bądź w połączeniu z analizą falkową (W) lub metodami bootstrapowymi (B) (tzn. ELM, ELMW, ELMB) do przewidywania dobrowolnego zapotrzebowania na wodę w mieście. Wyniki porównano z uzyskanymi tradycyjnymi metodami bazującymi na sztucznych sieciach neuronowych (tzn. ANN, ANNw, ANNb). Modele przewidyujące zapotrzebowanie na wodę zbudowano z wykorzystaniem trzyletniego zapotrzebowania wodę i zestawu danych klimatycznych dla miasta Calgary w kanadyjskiej prowincji Alberta. Hybrydowe modele ELMb i ANNb zapewniały satysfakcjonujące prognozy jednodniowe o podobnej dokładności, natomiast wyniki uzyskane z zastosowaniem modeli ELMW i ANNw były bardziej dokładne, przy czym model ELMW okazał się lepszy niż ANNw. Istotną poprawę prognozowania szczytowego zapotrzebowania na wodę w mieście uzyskano jedynie z zastosowaniem modelu ELMW. Wyższość modelu ELMW nad modelami ANNw czy ANNb dowodzi znaczącej roli transformacji falkowej w usprawnianiu działania modeli prognozujących zapotrzebowanie na wodę w mieście.