Urban Water Journal

ISSN: 1573-062X (Print) 1744-9006 (Online) Journal homepage: http://www.tandfonline.com/loi/nurw20

Using extreme learning machines for short-term urban water demand forecasting

Soukayna Mouatadid & Jan Adamowski

To cite this article: Soukayna Mouatadid & Jan Adamowski (2016): Using extreme learning machines for short-term urban water demand forecasting, Urban Water Journal, DOI: 10.1080/1573062X.2016.1236133

To link to this article: http://dx.doi.org/10.1080/1573062X.2016.1236133

Published online: 04 Oct 2016.

URBAN WATER JOURNAL, 2016
http://dx.doi.org/10.1080/1573062X.2016.1236133

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Using extreme learning machines for short-term urban water demand forecasting

Soukayna Mouatadid and Jan Adamowski

Department of Bioresource Engineering, McGill University, QC, Canada

ABSTRACT
This study explores the ability of various machine learning methods to improve the accuracy of urban water demand forecasting for the city of Montreal (Canada). Artificial Neural Network (ANN), Support Vector Regression (SVR) and Extreme Learning Machine (ELM) models, in addition to a traditional model (Multiple linear regression, MLR) were developed to forecast urban water demand at lead times of 1 and 3 days. The use of models based on ELM in water demand forecasting has not previously been explored in much detail. Models were based on different combinations of the main input variables (e.g., daily maximum temperature, daily total precipitation and daily water demand), for which data were available for Montreal, Canada between 1999 and 2010. Based on the squared coefficient of determination, the root mean square error and an examination of the residuals, ELM models provided greater accuracy than MLR, ANN or SVR models in forecasting Montreal urban water demand for 1 day and 3 days ahead, and can be considered a promising method for short-term urban water demand forecasting.

ARTICLE HISTORY
Received 8 September 2015
Accepted 16 August 2016

KEYWORDS
Water resources management; regression model; artificial neural network; support vector machine

1. Introduction

In many countries around the world, the rising severity of climate change coupled with population growth and economic development are leading to an increase in urban water demand (Gleick 2011). Canada, despite its illusion of fresh water abundance, is no exception to this situation (Adamowski *et al.* 2012). To help ensure dependable water availability and minimize peak water use, reliable urban water demand forecast models must be developed. Such forecasting models allow water utilities to make operational and strategic decisions, thereby improving water security and water consumption sustainability (Beal and Stewart 2011). In this context, accurate short-term water demand forecasting is an important component of optimizing water supply systems and implementing effective water demand management programs (Ghiassi *et al.* 2008).

In the case of Montreal, given the partial deterioration of the city's water supply distribution networks, accurate short-term water demand forecasts could help water managers develop a better understanding of the dynamics and underlying factors that affect water use, provide information on when peak day events are likely to occur, and analyse the benefits and costs of water conservation (Adamowski *et al.* 2012). Generally, variation in urban water demand forecasting is mainly affected by temperature, precipitation, and past water demand (Jain *et al.* 2001, Bougadis *et al.* 2005, Adamowski *et al.* 2012). In the literature, urban water demand has been forecasted using various methods, including pattern-based methods as well as linear statistical models and non-linear models (Maidment and Parzen 1984, Alvisi *et al.* 2007, Bennett *et al.* 2013). The literature review in this study focuses on data-driven and self-learning demand forecasting models.

Time series models are widely used for water demand forecasting purposes. In an early study, a daily municipal water consumption time series model, drawing on rainfall and air temperature, was developed based on past data signals, and applied to the forecasting of water consumption in nine American cities (Maidment 1985). Other studies have developed time series models to forecast daily municipal water demand in Washington, D.C. (Smith 1988), Melbourne, Australia (Zhou *et al.* 2000), and Hong Kong, China (Wong *et al.* 2010). Other examined methods include Artificial Neural Networks (ANN), which have been applied to short-term urban water demand forecasting to address nonlinearity (Jain and Ormsbee 2002). Subsequent research has shown that ANNs outperform linear regression techniques in urban water demand forecasting (Jain and Ormsbee 2002, Bougadis *et al.* 2005, Adamowski 2008, Sahoo *et al.* 2009, Adamowski *et al.* 2012). In addition to ANNs, Support Vector Machines (SVMs) are another machine learning technique that have been used in water demand forecasting. Future urban water demand forecasts for a city in South-Eastern Spain, from an SVR-based model, were found to be more accurate than those from ANN or other hybrid machine learning methods (Herrera *et al.* 2010). Similarly, SVR models developed for 24-hour water demand forecasting for a residential district in Berlin outperformed seasonal autoregressive models (Braun *et al.* 2014). In general, SVR has been widely applied for water demand forecasts and the effectiveness of this method has been verified (Jain and Ormsbee 2002, Msiza *et al.* 2007, Ahmadaali *et al.* 2013, Zhang *et al.* 2013).

CONTACT Soukayna Mouatadid soukayna.mouatadid@mail.mcgill.ca

Recently, the Extreme Learning Machine (ELM) method has been introduced in forecasting applications in computational science, but remains uncommon in hydrological forecasting research (Ravinesh and Şahin 2015). One study has shown that ELM has the potential to outperform ANN models in forecasting monsoon rainfall over south peninsular India (Acharya *et al.* 2013). An ELM model developed to forecast monthly Effective Drought Index, using input variables from meteorological datasets and climate mode indices, was found to significantly outperform an ANN model (Ravinesh and Şahin 2015). In another study, a model combining wavelet neural networks with ELM (WNN_ELM) to forecast discharge from two reservoirs in South-western China at a one month lead time was found to provide more accurate forecasts than an SVM model (Li and Cheng 2014). A study of the use of an ensemble empirical mode decomposition based ELM model (EEMD-ELM) for landslide displacement forecasting found such a model to outperform single ANN forecasting models (Lian *et al.* 2014).

In this study, water demand for the city of Montreal was forecasted for one- and three-day lead times, using one linear forecasting method (MLR) and three non-linear, self-learning methods: ANN, SVR and ELM. The aim of the study was to determine which method results in the most accurate forecasts. In order to do so, a model using ELM as a learning method to forecast urban water demand was proposed and tested for one- and three-day lead times. The performance of the developed ELM model was then compared to that of the developed MLR, ANN and SVR models. The study explores, for the first time, the use of an ELM-based method to forecast short-term urban water demand.

2. Theoretical background

Due to space limitations, this section focuses on a brief theoretical background for ELM. The remaining methods (MLR, ANN, and SVR) are well known and only references are provided for the interested reader: MLR (Glantz and Slinker 2001, Yan and Xiaogang 2009), ANN (Bishop 1995, Haykin 1999, Cimen 2008), and SVR (Vapnik 1995, Smola 1996, Smola and Scholkopf 2004).

2.1. Extreme learning machine

Extreme Learning Machine (ELM), first proposed in 2006 (Huang *et al.* 2006b), is based on a Single Layer Feed-forward Neural Network (SLFN) architecture (Rajesh and Siva Prakash 2011). A typical SLFN structure consists of an input layer, a hidden layer and an output layer, which are connected by neurons (Zhang *et al.* 2014). This method is limited to feed-forward neural networks with a single nonlinear hidden layer, as this type of network is capable of forming an arbitrarily close approximation of any continuous nonlinear mapping (Fei *et al.* 2006). Instead of using traditional gradient-based learning methods which involve several iterations, ELM has its input weights randomly generated and uses simple matrix computations to determine the output weights (Chen *et al.* 2012, Wan *et al.* 2014). Compared with traditional training methods, ELM has much better generalization performance and requires less human intervention (Chen *et al.* 2012). Since it automatically determines all network parameters, ELM avoids trivial human intervention and proves to be efficient in online and real-time applications (Rajesh and Siva Prakash 2011). Moreover, it has much faster learning and running speed than more conventional machine learning methods (Huang *et al.* 2004). The mechanisms of ELM, as used in this study, can be described as follows.

Given N arbitrary distinct samples, $\{(v_i, t_i)\}_{i=1}^{N}$, where $v_i \in R^n$ with $v_i = [v_{i1}, v_{i2}, \ldots, v_{in}]^T$ and $t_i \in R^m$ with $t_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T$, ELM with K hidden nodes and an activation function $g(\cdot)$ can be mathematically modelled as (Wan *et al.* 2014):

$$f(v_j;u,r,\beta) = \sum_{i=1}^{K} \beta_i g(\boldsymbol{u}_i \cdot \boldsymbol{v}_j + r_i), \quad j = 1, \ldots, N \tag{1}$$

Where $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ is the weight vector connecting the ith hidden node and the output, $u_i = [u_{i1}, u_{i2}, \ldots, u_{in}]^T$ is the weight vector connecting the ith hidden node and the inputs, and r_i is the threshold of the ith hidden node.

If ELM can approximate the data samples with zero error, it means that (Wan *et al.* 2014):

$$\sum_{i=1}^{K} \beta_i g(\boldsymbol{u}_i \cdot \boldsymbol{v}_j + r_i) = \boldsymbol{t}_j \quad j = 1, \ldots, N \tag{2}$$

Equation (1) can then be written compactly as (Wan *et al.* 2014):

$$H\beta = T \tag{3}$$

$$H = \begin{bmatrix} g(u_1 \cdot v_1 + r_1) & \cdots & g(u_K \cdot v_L + r_K) \\ \vdots & \ddots & \vdots \\ g(u_1 \cdot v_N + r_1) & \cdots & g(u_K \cdot v_N + r_K) \end{bmatrix}_{N\times K} \tag{4}$$

Where $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{in}]^T$ denotes the matrix output weights, H is the hidden-layer output matrix of the ELM, and $T_i = [t_{i1}, t_{i2}, \ldots, t_{in}]^T$ denotes the matrix of targets. The input weights u_i and the hidden-layer biases r_i are randomly generated using continuous probability distributions and are not necessarily tuned. Once random values have been assigned to these parameters in the beginning of learning, matrix H can remain unchanged.

If the activation function is infinitely differentiable, when the number of hidden neurons is equal to the number of distinct training samples ($K = n$), the parameters of hidden nodes can be assigned randomly and the output weights can be analytically determined by simply inverting the matrix. Therefore, ELM can approximate data samples with zero error. If $K << n$, then H becomes a non-square matrix. One specific set of r_j can be determined such that:

$$\| H\left(\hat{u}_1, \ldots .\hat{u}_K, \hat{r}_1, \ldots .\hat{r}_K\right)\hat{\beta} - T \| = \min \| H\left(u_1, \ldots .u_K, r_1, \ldots .r_K\right)\hat{\beta} - T \| \tag{5}$$

This is equivalent to minimizing the cost function of the traditional gradient-based learning algorithms used in back-propagation learning (Wan *et al.* 2014):

$$C_{BP} = \sum_{j=1}^{N} \left[\sum_{i=1}^{K} \beta_i g\left(u_i \cdot v_j + r_i\right) - t_j \right]^2 \tag{6}$$

Since the input weights and the hidden-layer biases are randomly assigned and fixed, training an SLFN is then equivalent

Figure 1. Map showing the study location: Montreal, Canada (CGTSIM 2015).

to finding a least-squares solution for the linear system. For the given u_i, and r_i, Equation (2) becomes a linear system, and the smallest norm least-squares solution is:

$$\hat{\beta} = H^{\top} T \tag{7}$$

where $H^{\top}$ is the Moore–Penrose generalized inverse of matrix H. The singular value decomposition (SVD) method is almost universally used to obtain $H^{\top}$ (Huang *et al.* 2006a), and was hence deemed the most appropriate method to use in the present study.

2.2. Performance evaluation of various models

The two performance indices used to evaluate the performance of the developed MLR, ANN, SVR and ELM models were the coefficient of determination (R^2) and the root mean squared error ($RMSE$):

$$R^2 = 1 - \sum_{i=1}^{n} \left(O_i - P_i\right)^2 / \sum_{i=1}^{n} \left(O_i - \acute{O}\right)^2 \tag{8}$$

$$RMSE = \left((1/n) \sum_{i=1}^{n} \left(O_i - P_i\right)^2 \right)^{1/2} \tag{9}$$

Where $\acute{O}$ is the mean observed water demand, O_i is the observed water demand, P_i is the forecasted water demand, and n is the number of data points. In addition, the residuals of the best performing models from each method were examined. Four q-q plots were used to verify the normal distribution of forecasting errors, along with histograms of the frequency distribution of the forecasting errors and boxplots of the observation and forecast values.

3. Study area and data

Montreal is the second largest city in Canada, with a population of approximately 1.65 million individuals (Statistics Canada 2012). The city has six drinking water treatment plants, with a total production capacity of 2.9×10^6 m^3 d^{-1}. Figure 1 shows the study location.

The city's 5000 km water distribution system is old and suffers from pipe breaks and water leakage problems. Past studies have shown that approximately 33% of the treated water is lost (City of Montreal 2014), hence the need to repair water lines and improve water leak detection techniques (Tiwari and Adamowski 2013). Since 2002, the city has begun major renovation efforts and repairs to the distribution system. It has been estimated that this will take a minimum of 20 years and cost several billion dollars (City of Montreal 2010). On average, Montreal produces 934 L d^{-1} of drinking water per capita (City of Montreal 2014).

The data used in the study were obtained from the City of Montreal and Environment Canada and consisted of average daily water demand (D), maximum temperature (T_{max}), total precipitation (R), and occurrence of precipitation (CR) from 27 February 1999 to 6 August 2010. Additional data were not available. The data used to develop the MLR models was partitioned into two sets: the first set contained 80% of the data and was used to train the models, the second set contained 20% of the data and was used to test the models. For the development of the ANN models, it is common practice to divide the available data into three sub-sets: a training, validation, and testing set (Maier and Dandy 2000). In this case, approximately 80% of the data was used to train the models, 10% was used for cross-validation to check that the models do not over fit, and the remaining 10% was used for testing the performance of the developed models. Regarding the

SVR and ELM models, 90% of the data was used for calibrating the models and the remaining 10% was used for testing the models.

4. Model development

Water demand forecasting can be based on different time scales, depending on the purpose for which the model is developed (Bakker *et al.* 2003). When it comes to short-term forecasting (one day to one week), these types of models can generally be used for the daily operations of distribution systems or treatment plants (Bakker *et al.* 2013). In this study, a short-term scale was adopted and the water demand was forecasted for one day and three days ahead, i.e., for *d*+1 and *d*+3 days' lead time, respectively.

This section describes how the urban water demand was forecasted over two separate lead times using MLR, ANN, SVR and ELM models. For each machine learning method used in this study, optimal model parameters were chosen such that they optimize the coefficient of determination R^2. In addition, the inputs were standardized to fall within a range of [0, 1]. By standardizing the variables and recasting them into dimensionless units, the arbitrary effect of similarity between objects is removed (Sudheer *et al.* 2002). These variables at the current day (*d*) as well as the corresponding delayed variables (*d*–1, *d*–2, etc.) were used as input variables. In order to determine the degree of delay in the variables, a trial and error procedure was adopted where the delay in the lagged variables was continuously increased, starting from (*d*–1) until (*d*–4) when the performance of the models started to deteriorate. Consequently, the following variables were used as inputs: the daily urban water demand (*D*), maximum temperature (T_{max}), total precipitation (*R*) and the occurrence of precipitation (*CR*) recorded at the current day (*d*), one day ago (*d*–1), two days ago (*d*–2) and three days ago (*d*–3), in order to forecast the water demand at (*d*+1) and (*d*+3).

Once the delayed variables to be used were determined, the next step was to identify the optimal combination of input variables which leads to the highest forecasting accuracy. Since there is no direct method to do so, a trial and error approach was used, which is based on feeding the models different combinations of the input variables. The current day variables and delayed variables were grouped into different combinations. The first combination to be tested included all four input variables, then one input variable was removed at a time (e.g., all variables except the maximum temperature), then two at a time, and then three at a time. The tests were repeated for the case where the variables used were from the current day (*d*) until (*d*–3), from the current day (*d*) until (*d*–2), from the current day (*d*) until (*d*–1), and finally only from the current day (*d*).

Then, for each learning method, the models' architecture was tested by varying the parameters' values specific to each method. For ANN, the combination of input variables and parameters that produced the lowest generalization error was identified as the optimal structure. For MLR, SVR and ELM, the combination of input variables and parameters that produced the highest R^2 values for the training data sets was selected as the optimum structure. The same procedure was used for the one day and three days ahead forecasts.

4.1. MLR model development

In this study, all MLR models were developed using Microsoft Excel 2013 software. Urban water demand estimates were set as the dependent *y* variable and the maximum temperature, precipitation, the occurrence of rain and the urban water demand were set as the independent *x* variables. A trial and error method was used to determine what combination of input variables produced the highest R^2 value. The performance of each model was compared using statistical measures of fit.

4.2. ANN model development

In this study, the ANN models had a feed-forward multi-layer perceptron architecture and were trained using a Levenberg-Marquardt (LM) back propagation algorithm. The activation function used for the hidden layer was the 'logsig' function, whereas the 'purelin' function was used for the output layer. For each ANN model, there were between three to six inputs. The optimal number of hidden neurons is usually determined empirically as being equal to log (*N*), where *N* is the number of training samples (Wanas *et al.* 1998); however, another study has shown that the best performance of a neural network occurs when the number of hidden neurons is $2n + 1$, where *n* is the number of input neurons (Mishra and Desai 2006). These two methods were used to establish an upper and lower bound of fifteen and three for the number of hidden neurons in this study. The optimal number of hidden neurons was determined by trial and error as the number of neurons which produced the lowest generalization error (Jia and Culver 2006). All ANN models were created using the Neural Network Toolbox of MATLAB.

4.3. SVR model development

All SVR models were developed using the LibSVM library, which is designed to build SVMs for regression (Chang and Lin 2011). As all SVR models used the nonlinear radial basis function (RBF) kernel, each SVR model consisted of three parameters: gamma (γ), cost (*C*), and epsilon (ε). Gamma (γ) is a constant that reduces the model space and controls the complexity of the solution, *C* is a positive constant that is a capacity control parameter, while ε is the loss function that describes the regression vector without all the input data (Kisi and Cimen 2011). The values of γ, *C and* ε were varied between 0.2 and 0.47, 0.01 and 0.04, and 0.1 and 0.4, respectively. The selection of the input variables was based on a trial and error approach. The combination of input variables and parameters that produced the highest R^2 values for the training data sets was selected as the optimum SVR structure (Belayneh and Adamowski 2012).

4.4. ELM model development

The ELM models were all developed in the MATLAB 2014a environment (Huang *et al.* 2012). In designing the ELM model, three layers were used to build the architecture for forecasting the water demand. The first layer had a number of neurons varying from one to sixteen, whereas the output layer had one neuron representing the forecasted water demand. The hidden layer was tested with a lower limit of 10 hidden neurons and an upper limit of 100 hidden neurons. Different activation functions were tested one by one, including the sigmoid 'Sig', the sine 'Sin', the hard-limit 'Hardlim', the triangular basis 'Tribas' and the radial basis function 'Radbas'. For each trial, the number of nodes in the

Table 1. Results of the best models for the four compared methods during training, validation and testing.

Model	Input combination	Model structure	Train R^2	Train *RMSE*	Validate R^2	Validate *RMSE*	Test R^2	Test *RMSE*
1 Day Lead Time								
ELM	$D_d, D_{d-1}, D_{d-2}, R_d, R_{d-1}, R_{d-2}$	(6-60-1) - 'sig'	0.87683	9.50499	–	–	0.86847	7.66878
SVR	$D_d, D_{d-1}, D_{d-2}, D_{d-3}, R_d, R_{d-1}, R_{d-2}, R_{d-3}$	(8 - 1) - (0.2)	0.84943	10.50353	–	–	0.84930	8.20423
ANN	$D_d, D_{d-1}, D_{d-2}, D_{d-3}$, $Tmax_d$, $Tmax_{d-1}$, $Tmax_{d-2}$, $Tmax_{d-3}$, CR_d, CR_{d-1}, CR_{d-2}, CR_{d-3}	(12 -11-1)	0.85565	10.28451	0.82851	7.80062	0.82812	9.25469
MLR	$D_d, D_{d-1}, D_{d-2}, D_{d-3}, R_d, R_{d-1}, R_{d-2}, R_{d-3}$	(8-1)	0.85368	10.35431	–	–	0.85710	7.98920
3 Days Lead Time								
ELM	$D_d, D_{d-1}, D_{d-2}, D_{d-3}$, $Tmax_d$, $Tmax_{d-1}$, $Tmax_{d-2}$, $Tmax_{d-3}$	(8-90-1) - 'sin'	0.7742	12.8731	–	–	0.72900	11.01959
SVR	$D_d, D_{d-1}, D_{d-2}, D_{d-3}, R_d, R_{d-1}, R_{d-2}, R_{d-3}$	(8 - 1) - (0.2)	0.72394	14.2264	–	–	0.71792	11.23653
ANN	$D_d, D_{d-1}, D_{d-2}, D_{d-3}$, $Tmax_d$, $Tmax_{d-1}$, $Tmax_{d-2}$, $Tmax_{d-3}$	(8-12-1)	0.74846	13.57986	0.6896	10.49005	0.68333	12.59636
MLR	$D_d, D_{d-1}, D_{d-2}, D_{d-3}, CR_d, CR_{d-1}, CR_{d-2}, CR_{d-3}$	(8-1)	0.72601	14.17294	–	–	0.68488	11.87638

Table 2. Results of the best ELM models during training and testing.

Model	Input combination	Model structure	Train R^2	Train *RMSE*	Test R^2	Test *RMSE*
1 Day Lead Time						
ELM	$D_d, D_{d-1}, D_{d-2}, R_d, R_{d-1}, R_{d-2}$	(6-60-1) - 'sig'	0.87683	9.50499	0.86847	7.66878
	$D_d, D_{d-1}, D_{d-2}, D_{d-3}, R_d, R_{d-1}, R_{d-2}, R_{d-3}$	(8-70-1) - 'sig'	0.87753	9.47792	0.86387	7.80182
	$D_d, D_{d-1}, D_{d-2}, R_d, R_{d-1}, R_{d-2}$	(6-100-1) - 'sig'	0.88176	9.31310	0.86322	7.82047
	$D_d, D_{d-1}, D_{d-2}, R_d, R_{d-1}, R_{d-2}$	(6-90-1) - 'sig'	0.88039	9.36689	0.86301	7.82633
	$D_d, D_{d-1}, D_{d-2}, D_{d-3}, R_d, R_{d-1}, R_{d-2}, R_{d-3}$	(8-80-1) - 'sig'	0.87910	9.41716	0.86107	7.88159
3 Days Lead Time						
ELM	$D_d, D_{d-1}, D_{d-2}, D_{d-3}$, $Tmax_d$, $Tmax_{d-1}$, $Tmax_{d-2}$, $Tmax_{d-3}$	(8-90-1) - 'sin'	0.77420	12.87310	0.72900	11.01959
	$D_d, D_{d-1}, D_{d-2}, D_{d-3}$, $Tmax_d$, $Tmax_{d-1}$, $Tmax_{d-2}$, $Tmax_{d-3}$	(8-90-1) - 'sig'	0.77585	12.82600	0.72031	11.19478
	$D_d, D_{d-1}, D_{d-2}, D_{d-3}, CR_d, CR_{d-1}, CR_{d-2}, CR_{d-3}$	(8-70-1) - 'sin'	0.77306	12.90568	0.71177	11.36441
	$D_d, D_{d-1}, D_{d-2}, D_{d-3}, R_d, R_{d-1}, R_{d-2}, R_{d-3}, CR_d, CR_{d-1}, CR_{d-2}$	(11-50-1) - 'sig'	0.72995	14.07806	0.71019	11.39553
	$D_d, D_{d-1}, D_{d-2}, D_{d-3}, R_d, R_{d-1}, R_{d-2}, R_{d-3}, CR_d, CR_{d-1}, CR_{d-2}$	(11-60-1) - 'sin'	0.72913	14.09953	0.70990	11.40124

hidden layers was incremented by 10. Near optimal ELM models were obtained when the sigmoid function was used as the activation function.

5. Results and discussion

The training and testing phase accuracy, as quantified by the R^2 and *RMSE*, is presented for the best MLR, ANN, SVR, and ELM models in (Table 1), while the performance of the best 1- and 3-day lead time ELM models is presented in (Table 2).

In these tables, the structure of the ELM and ANN models is presented as: 'Number of input neurons - Number of hidden neurons - Number of output neurons'. The activation function of each ELM model is presented as well. When it comes to SVR and MLR models, the structure is written as 'Number of input neurons - Number of Output neurons'. The value gamma is presented between brackets for each SVR model.

Compared to MLR, ANN and SVR methods, ELM provided greater accuracy for 1- and 3-day lead forecasts of urban water demand for the city of Montreal. Generally, all data-driven models showed better results for forecasts of 1-day lead time than forecasts of 3-days lead time. As the lead-time was increased, all the models deteriorated in performance, with respect to both R^2 and *RMSE*. Nonetheless, this deterioration is not drastic and therefore does not result in poor models. For example, for the ELM models, the training R^2 and testing R^2 deteriorate by 11.7% and 16.06%, respectively. The small percentages indicate that the data-driven models developed can be considered reliable for forecasting urban water demand for 1- and 3-day lead times.

Figure 2 shows the spread of the observed and forecasted water demand values using a boxplot for all four methods. The horizontal lines extending from the top and bottom of the boxplots represent the whiskers, with the lower whisker representing the smallest non-outlier in the dataset and the upper whisker representing the largest non-outlier in the data set. Figure 2 shows that, overall, all four methods lead to a quartile distribution similar to that of the observation dataset, but the top and bottom whisker are slightly different for each forecasted water demand. Also, the median of the forecasts and observations is relatively similar for all four methods. The higher performance of the ELM models is confirmed by the smaller shift between the observation and the ELM forecasts.

The q-q plots in Figure 3 show that not all methods lead to normally distributed forecast errors. The forecasting errors approach more of a normal distribution for the 3-day lead time forecasts, compared to the 1-day lead time forecasts. In the case of ELM, the q-q plot forecast errors of the 1-day ahead forecast presents a left skew and the forecast errors of the 3-day ahead forecast present a more normal q-q plot.

Figure 4 illustrates the histogram of frequency distribution of the forecasting error (PE) for the testing dataset. The percentages of under-forecasting (PE < 0) and over-forecasting (PE > 0) for each method are also shown. It was found that while MLR forecasting errors are mainly over-forecasts, the forecasting errors for the other methods are either over-forecasts or under-forecasts.

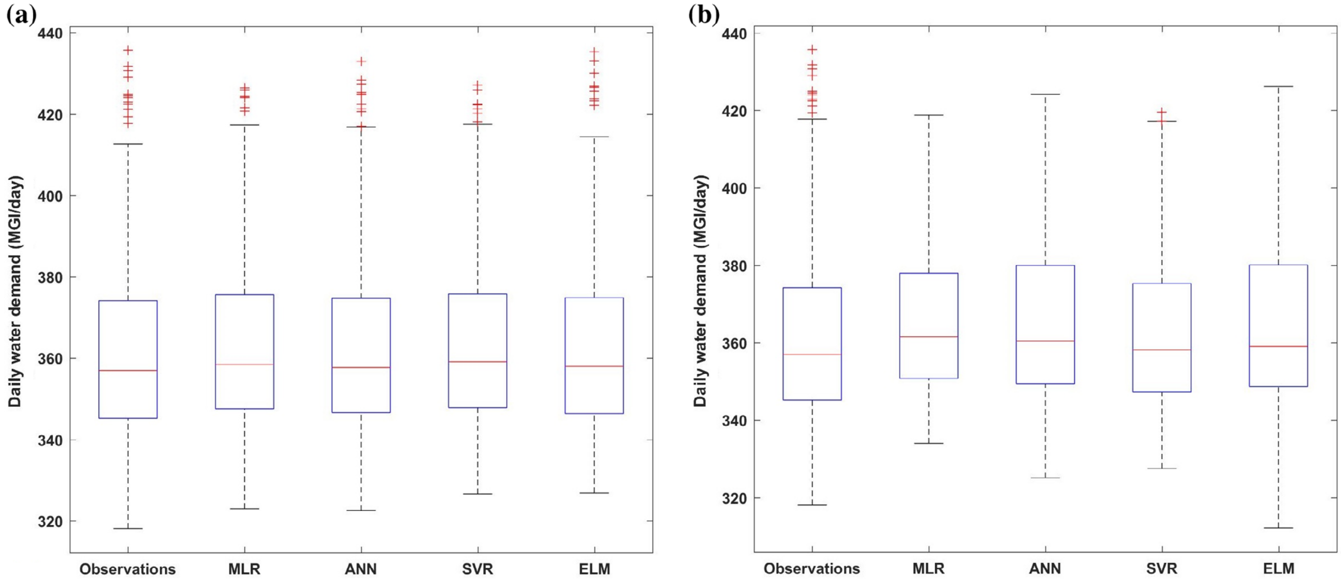

Figure 2. Boxplots of daily water demand values (MGl/day) from observed and forecasted data using MLR and ANN models, calculated for the test period, for one day ahead (a) and three days ahead (b).

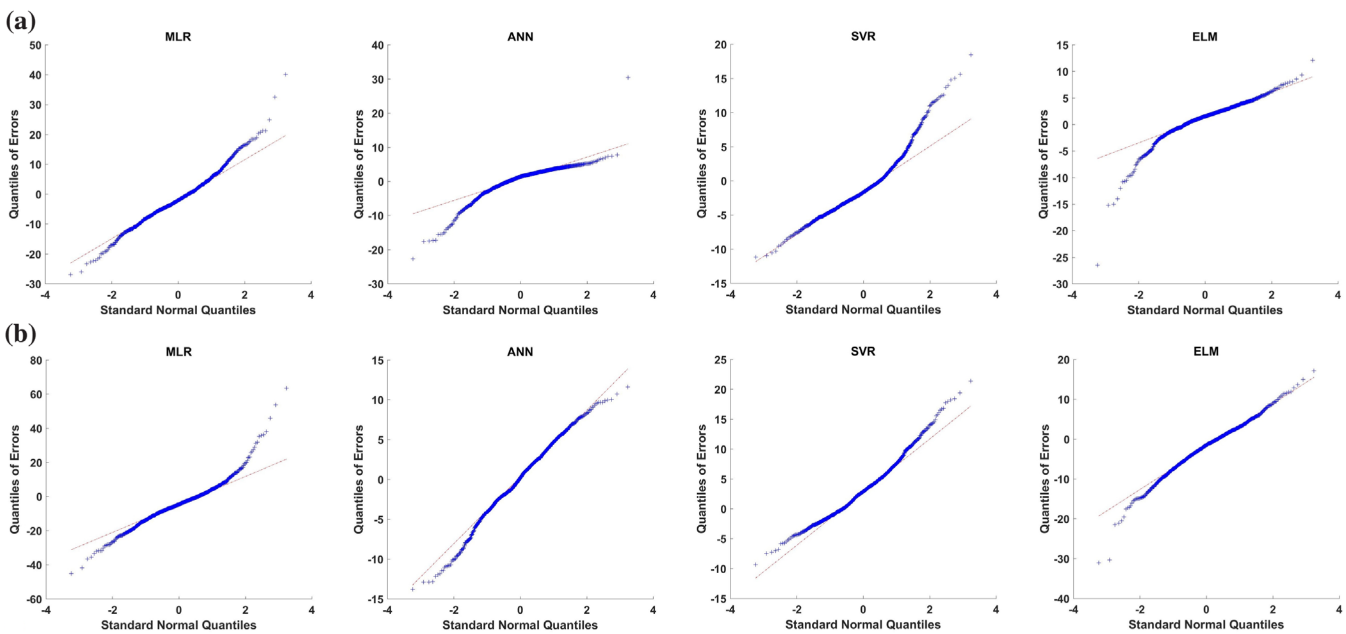

Figure 3. Q-Q plots of the MLR, ANN, SVR and ELM forecasting error (PE) calculated for the test period, for one day ahead (a) and three days ahead (b).

For ELM for example, the majority of forecast errors are under-forecasts for the 1-day ahead forecasts and over-forecasts for the 3-day ahead forecasts.

In addition, ELM as a learning method outperforms MLR, ANN and SVR in terms of learning speed (the total training and testing time for a given set of samples). For the MLR, ANN and SVR models, the total running time increases with the number of training samples, from less than a minute for a hundred samples to approximately five minutes for a thousand samples. For ELM modelling, however, the running time is shorter (less than one second) and does not increase with the addition of more training samples. In fact, for a sample size of 4173 entries, the learning speed was 0.35 seconds for ELM and 4.5 minutes on average for MLR, ANN and SVR. In other words, the learning speed for ELM increased by 99.8% compared to the other three methods discussed.

Also, it appears that the network architecture of the model influences the forecasting performance for the ELM models. From (Table 2), it is clear that the use of the sigmoid and the sine functions, as activation functions for the ELM model, leads to more accurate results, as opposed to the other tested activation functions, namely the hard-limit, the triangular basis and the radial basis functions. In general, ELM is an efficient learning method when it comes to short-term forecasting of urban water demand forecasting.

One possible reason for the better performance of the ELM models could be the highly efficient tuning mechanism of ELM,

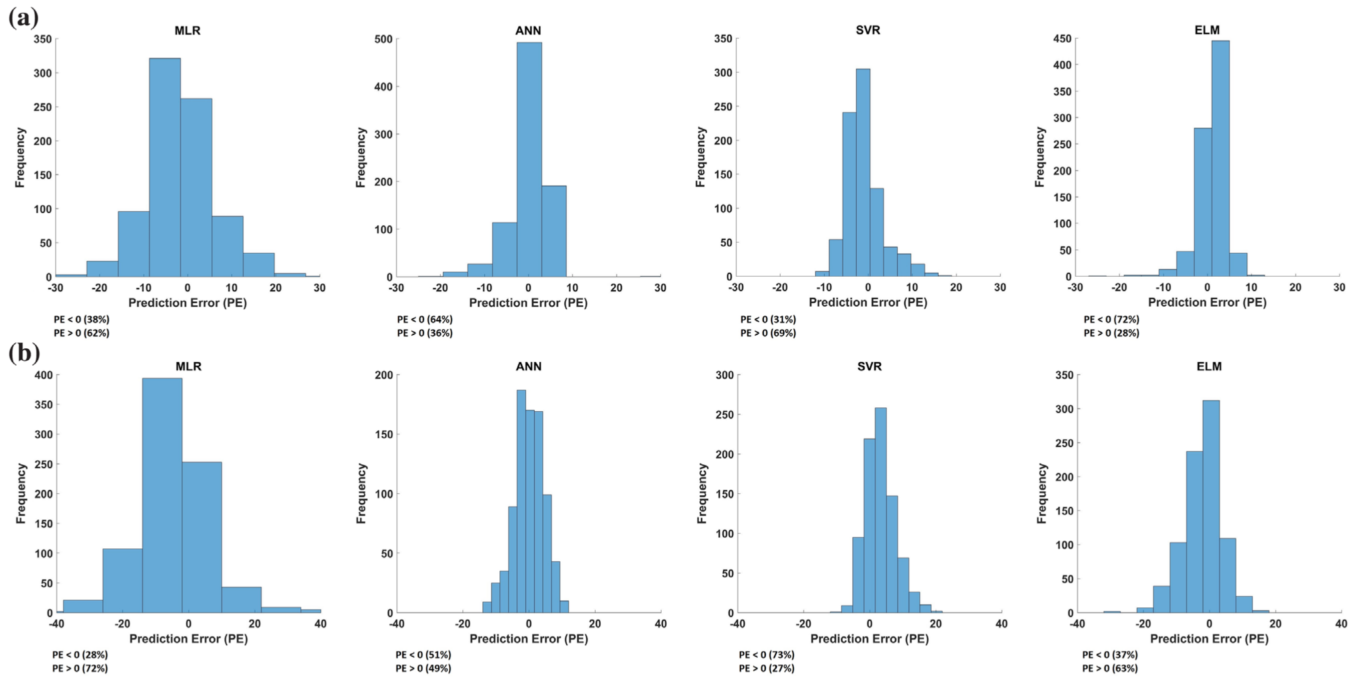

Figure 4. Histogram of frequency distribution of the forecasting error (PE) calculated for the test period, for one day ahead (a) and three days ahead (b). The cumulative frequency represents under-forecasts (PE < 0) and over-forecasts (PE > 0) by MLR, ANN, SVR and ELM models.

which significantly improves the forecasting performance (Chen *et al.* 2012). Moreover, ELM presents other advantages over the other three studied methods, as it saves running time and requires fewer user-defined parameters (Şahin *et al.* 2014). Therefore, ELM can considerably reduce computational complexity in comparison to the other three methods, which is useful for large-scale data analysis and forecasting applications. In contrast, SVM presents two disadvantages. Training time scales can be between quadratic and cubic depending on the number of training samples, hence increasing as the size of the training set increases. The selection of a kernel function and its specific parameters can also be difficult (Adamowski and Sun 2010).

It should be noted that, since the forecasting models presented in this study did not account for uncertainty in the data, the reader should consider the error in forecasts which generally results from data uncertainty. Also, the comparisons discussed in this study are based on the global performance of the developed models, but do not consider the time varying performance of these models. A comparative study which investigates whether some methods perform better than others at some points in the year would be interesting. And finally, since the combination of wavelet transforms and ANNs and SVMs have provided good forecasting results in hydrology (Kisi and Cimen 2011), a comparative study between WA-ANN, WA-SVM, and WA-ELM applied to short-term urban water demand forecasting would be useful.

6. Conclusions and recommendations

Accurate and reliable urban water demand forecasting is a necessary pre-requisite for effective and sustainable urban water resources planning and management. This study investigated the ability of data driven models to forecast urban water demand. The study was carried out to develop an urban water demand forecasting model that is accurate and reliable for 1- and 3-day lead time urban water demand forecasting. To do so, the potential of MLR, ANN, SVR and ELM models was investigated; the study site was the City of Montreal, Canada. Overall, ELM models were found to provide better results than the other model types used to forecast with one- and three-day lead times.

The superior performance of the ELM model, when compared with the ANN and SVR models for one-day and three-day lead times, demonstrates the usefulness of using a random weights assignment. The best ELM models for 1-day and 3-day lead time water demand forecasting considered water demand, total precipitation and maximum temperature to play a role in improving the model's accuracy. This implies that all the explanatory variables used in this study are needed for urban water demand modelling.

As indicated by the results of this study, ELM is a promising new forecasting method that can improve the accuracy of short-term urban water demand forecasts over the MLR, ANN and SVR methods. Therefore, models developed using ELM can be used in similar applications which require thousands of samples to be processed in a time efficient manner. Possible future studies could investigate which of these data driven models is suitable for forecasting long-term water demand values in other locations with different physical characteristics and different climates. Additionally, further research could determine whether there is a significant link between forecast accuracy and climate. ELM could also be tested for other hydrological applications such as groundwater level forecasting, streamflow forecasting, and rainfall forecasting.

Funding

This research was funded by an NSERC Discovery Grant held by Jan Adamowski.

References

Acharya, N., Shrivastava, N.A., Panigrahi, B. and Mohanty, U., 2013. Development of an artificial neural network based multi-model ensemble to estimate the northeast monsoon rainfall over south peninsular India: An application of extreme learning machine. *Climate Dynamics*, 43, 1303–1310.

Adamowski, J., 2008. Peak daily water demand forecast modeling using artificial neural networks. *Journal of Water Resources Planning and Management*, 134 (2), 119–128.

Adamowski, J. and Sun, K., 2010. Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology*, 390 (1–2), 85–91.

Adamowski, J., Fung Chan, H., Prasher, S.O., Ozga-Zielinski, B. and Sliusarieva, A., 2012. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resources Research*, 48, W01528.

Ahmadaali, K., Liaghat, A., Haddad, O.B. and Heydari, N., 2013. Estimation of virtual water using support vector machine, K-nearest neighbour, and radial basis function neural network models. *International Journal of Agronomy and Plant Production*, 4 (11), 2926–2936.

Alvisi, S., Franchini, M. and Marinelli, A., 2007. A short-term, pattern-based model for water-demand forecasting. *Journal of Hydroinformatics*, 9 (1), 39–50.

Bakker, M., Schagen, K.V. and Timmer, J., 2003. Flow control by prediction of water demand. *Journal of Water Supply: Research & Technology*, 52, 417–424.

Bakker, M., Vreeburg, J.H.G., Van, S.K.M. and Rietveld, L.C., 2013. A fully adaptive forecasting model for short-term drinking water demand. *Environmental Modelling & Software*, 48 (1), 141–151.

Beal, C. and Stewart, R.A., 2011. South East Queensland residential end use study. *Urban Water Security Research Alliance*, 47, 1836–5558.

Belayneh, A. and Adamowski, J., 2012. Standard precipitation index drought forecasting using neural networks, wavelet neural networks, and support vector regression. *Applied Computational Intelligence and Soft Computing*, 794061, 13.

Bennett, C., Stewart, R.A. and Beal, C.D., 2013. ANN-based residential water end-use demand forecasting model. *Expert Systems with Applications*, 40 (4), 1014–1023.

Bishop, C.M., 1995. *Neural networks for pattern recognition*. Oxford: Clarendon Press.

Bougadis, J., Adamowski, K. and Diduch, R., 2005. Short-term municipal water demand forecasting. *Hydrological Processes*, 19, 137–148.

Braun, M., Bernard, T., Piller, O. and Sedehizade, F., 2014. 24-hours demand forecasting based on SARIMA and support vector machines. *Procedia Engineering*, 89 (3), 926–933.

CGTSIM, 2015. *CGTSIM Annual report 2014-2015*. Montreal: Comité de gestion de la taxe scolaire de l'île de Montréal Presse.

Chang, C., and Lin, C., 2011. *LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27. Software Available at http://www.csie.ntu.edu.tw/~cjlin/libsvm/faq.html#f203

Chen, X., Dong, Z.Y., Meng, K., Xu, Y., Wong, K.P. and Ngan, H.W., 2012. Electricity price forecasting with extreme learning machine and bootstrapping. *IEEE Transactions on Power Systems*, 27 (4), 2055–2062.

Cimen, M., 2008. Estimation of daily suspended sediments using support vector machines. *Hydrological Sciences Journal*, 53 (3), 656–666.

City of Montreal, 2010. *The Montreal community sustainable development plan 2010–2015*. Montreal: The City of Montreal.

City of Montreal, 2014. *Bilan de l'Usage de l'Eau Potable 2013*. Montreal: The City of Montreal.

Fei, H., Huang, D.S., Zhu, Z.H. and Rong, T.H., 2006. The forecast of the postoperative survival time of patients suffered from non-small cell lung cancer based on PCA and extreme learning machine. *International Journal of Neural Systems*, 16 (1), 39–46.

Ghiassi, M., Zimbra, D. and Saidane, H., 2008. Urban water demand forecasting with a dynamic artificial neural network model. *Journal of Water Resources Planning and Management*, 134 (2), 138–146.

Glantz, S.A. and Slinker, B.K., 2001. *Primer of applied regression and analysis of variance*. 2nd ed. New York, NY: McGraw-Hill.

Gleick, P.H., 2011. *The world's water volume 7: The biennial report on freshwater resources*. Washington, DC: Island Press.

Haykin, S., 1999. *Neural networks: A comprehensive foundation*. Upper Saddle River, NJ: Prentice Hall.

Herrera, M., Torgo, L., Izquierdo, J. and Perez-Garcia, R., 2010. Predictive models for forecasting hourly urban water demand. *Journal of Hydrology*, 387 (1–2), 141–150.

Huang, G.B., Zhu, Q.Y., and Siew, C.K., 2004. Extreme learning machine: A new learning scheme of feedforward neural networks. *In: 2004 IEEE International Joint Conference on Neural Networks*, 25–29 July 2004, Budapest. Budapest: IEEE, 985–990.

Huang, G.B., Zhu, Q.Y. and Siew, C.K., 2006a. Extreme learning machine: Theory and applications. *Neurocomputing*, 70 (1), 489–501.

Huang, G.B., Chen, L. and Siew, C.K., 2006b. Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Transactions on Neural Networks*, 17 (4), 879–892.

Huang, G.B., Zhou, H., Ding, X. and Zhang, R., 2012. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, 42 (2), 513–529.

Jain, A., Varshney, K. and Joshi, U.C., 2001. Short-term water demand forecast modeling at IIT Kanpur using artificial neural networks. *Water Resources Management*, 15 (5), 299–321.

Jain, A. and Ormsbee, L.E., 2002. Short-term water demand forecast modeling techniques - Conventional methods versus AI. *American Water Works Association Journal*, 94 (7), 64–72.

Jia, Y. and Culver, T.B., 2006. Bootstrapped artificial neural networks for synthetic flow generation with a small data sample. *Journal of Hydrology*, 331 (3–4), 580–590.

Kisi, O. and Cimen, M., 2011. A wavelet-support vector machine conjunction model for monthly streamflow forecasting. *Journal of Hydrology*, 399 (1–2), 132–140.

Li, B.J. and Cheng, C.T., 2014. Monthly discharge forecasting using wavelet neural networks with extreme learning machine. *Science China Technological Sciences*, 57 (12), 2441–2452.

Lian, C., Zeng, Z., Yao, W. and Tang, H., 2014. Extreme learning machine for the displacement prediction of landslide under rainfall and reservoir level. *Stochastic Environmental Research and Risk Assessment*, 28 (8), 1957–1972.

Maidment, D. and Parzen, E., 1984. Monthly water use and its relationship to climatic variables in Texas. *Water Resources Bulletin*, 19 (8), 409–418.

Maidment, D., Miaou, S. and Crawford, M., 1985. Transfer function models of daily urban water use. *Water Resources Research*, 21 (4), 425–432.

Maier, H.R. and Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and application. *Environmental Modelling and Software*, 15 (1), 101–124.

Mishra, A.K. and Desai, I.V.R., 2006. Drought forecasting using feed-forward recursive neural network. *Ecological Modelling*, 198 (1–2), 127–138.

Msiza, I.S., Nelwamondo, F.V., and Marwala, T., 2007. Artificial neural networks and support vector machines for water demand time series forecasting. *In: ISIC. IEEE International Conference on Systems, Man and Cybernetics*, 7–10 October 2007, Montreal. Montreal: IEEE, 7–10.

Rajesh, R. and Siva Prakash, J., 2011. Extreme learning machines – A review and state-of-the-art. *International Journal of Wisdom Based Computing*, 1 (1), 35–39.

Ravinesh, C.D. and Şahin, M., 2015. Application of the extreme learning machine algorithm for the prediction of monthly effective drought index in eastern Australia. *Atmospheric Research*, 153, 512–525.

Şahin, M., Kaya, Y., Uyar, M. and Yldrm, S., 2014. Application of extreme learning machine for estimating solar radiation from satellite data. *International Journal of Energy Research*, 38 (2), 205–212.

Sahoo, G.B., Schladow, S.G. and Reuter, J.E., 2009. Forecasting stream water temperature using regression analysis, artificial neural network, and chaotic non-linear dynamic models. *Journal of Hydrology*, 378, 325–342.

Smola, A.J., 1996. *Regression estimation with support vector learning machines*. Munich: Technical University of Munich.

Smola, A.J. and Scholkopf, B., 2004. A tutorial on support vector regression. *Statistics and Computing*, 14 (3), 199–222.

Statistics Canada, 2012. *Focus on Geography Series, 2011 Census*. Ottawa: Statistics Canada.

Sudheer, K.P., Gosain, A.K., Rangan, D.M. and Saheb, S.M., 2002. Modelling evaporation using an artificial neural network algorithm. *Hydrological Processes*, 16 (16), 3189–3202.

Tiwari, M.K. and Adamowski, J.F., 2013. Urban water demand forecasting and uncertainty assessment using ensemble wavelet–bootstrap neural network models. *Water Resources Research*, 49 (10), 6486–6507.

Vapnik, V., 1995. *The nature of statistical learning theory*. New York, NY: Springer.

Wan, C., Xu, Z., Xu, Z., Pinson, P., Dong, Z.Y. and Wong, K.P., 2014. Probabilistic forecasting of wind power generation using extreme learning machine. *IEEE Transactions on Power Systems*, 29 (3), 1033–1044.

Wanas, N., Auda, G., Kamel, M.S. and Karray, F., 1998. On the optimal number of hidden nodes in a neural network. *In: Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering*, 24–28 May 1998, Waterloo. Ontario. Waterloo: IEEE, 918–921.

Wong, J.S., Zhang, Q. and Chen, Y.D., 2010. Statistical modeling of daily urban water consumption in Hong Kong: Trend, changing patterns, and forecast. *Water Resources Research*, 46 (3), W03506.

Yan, X. and Xiaogang, S., 2009. *Linear regression analysis theory and computing*. Singapore: World Scientific Publishing.

Zhang, Q., Diao, Y. and Dong, J., 2013. Regional water demand prediction and analysis based on Cobb-Douglas model. *Water Resources Management*, 27 (8), 3103–3113.

Zhang, W., Zhao, W. and Du, X., 2014. Short-term forecast technology in load of electrified railway based on wavelet-extreme learning machine. *Journal of Networks*, 9 (6), 1502–1508.

Zhou, S., McMahon, T., Walton, A. and Lewis, J., 2000. Forecasting daily urban water demand: A case study of Melbourne. *Journal of Hydrology*, 236 (3), 153–164.