

Annals of Warsaw University of Life Sciences – SGGW
 Land Reclamation No 46 (3), 2014: 197–204
 (Ann. Warsaw Univ. of Life Sci. – SGGW, Land Reclam. 46 (3), 2014)

Weekly urban water demand forecasting using a hybrid wavelet–bootstrap–artificial neural network approach

KAZ ADAMOWSKI¹, JAN F. ADAMOWSKI², OUSMANE SEIDOU¹,
 BOGDAN OZGA-ZIELIŃSKI³

¹Department of Civil Engineering, University of Ottawa

²Department of Bioresource Engineering, McGill University

³Department of Environment Protection and Development, Warsaw University of Technology

Abstract: *Weekly urban water demand forecasting using a hybrid wavelet–bootstrap–artificial neural network approach.* This study developed a hybrid wavelet–bootstrap–artificial neural network (WBANN) model for weekly (one week) urban water demand forecasting in situations with limited data availability. The proposed WBANN method is aimed at improving the accuracy and reliability of water demand forecasting. Daily maximum temperature, total precipitation and water demand data for almost three years were used in this study. It was concluded that the hybrid WBANN model was more accurate compared to the ANN, BANN and WANN methods, and can be applied successfully for operational water demand forecasting. The WBANN model simulated peak water demand very effectively. The better performance of the WBANN model indicated that wavelet analysis significantly improved the model's performance, whereas the bootstrap technique improved the reliability of forecasts by producing ensemble forecasts. The WBANN model was also found to be effective in assessing the uncertainty associated with water demand forecasts in terms of confidence bands; this can be helpful in operational water demand forecasting.

Key words: artificial neural networks, resampling, ensemble, water demand, forecasting, wavelets, Canada

INTRODUCTION

The efficient operation and management of an urban water supply system requires accurate water demand forecasts, and the

estimation of future urban water demand is critical to the sustainable planning of water supply systems. This study focuses on urban water demand forecasts for one week lead times. Traditional or conventional urban water demand forecasting has used linear regression, trend-extrapolation, and time series. Adamowski and Karapataki (2010) and Adamowski et al. (2012) found that ANN models performed better than multiple linear regression models for peak weekly water demand forecasts. Adamowski et al. (2013) investigated the use of the continuous wavelet transform to determine changes in the temporal pattern of urban water demand and its potential meteorological drivers, and found that in areas with low precipitation (e.g. Calgary, Canada), there was an inverse relationship between urban water demand and precipitation during the summer months. One major drawback of ANNs, however, has been their inability to address data non-stationarity. Wavelet analysis has been found to be a potentially useful method for detecting non-stationarity (Adamowski et al. 2013). To address this issue, a hybrid wavelet–bootstrap–artificial neural network (WBANN) is explored in this paper. The hybrid wave-

let-bootstrap-ANN (WBANN) model was developed and compared with the hybrid bootstrap-based ANN (BANN) and hybrid wavelet-based ANN (WANN) methods to assess the effectiveness of the models.

METHODOLOGY

Artificial neural networks (ANN)

The multilayer feed-forward neural network developed in this study consists of an input layer, one or more hidden layers of computation nodes, and an output layer of computation nodes. An output node of an ANN is calculated as:

$$O_k = g_z \left[\sum_j V_j w_{ij} g_1 \left(\sum_i w_{ji} l_i + w_{j0} \right) + w_{k0} \right] \quad (1)$$

where:

O_k – output at the k^{th} node through the activation function g_2 ,

w_{ji} – adjustable weight connecting the i^{th} input node and the j^{th} hidden node of the respective input and hidden layers,

g_1 – activation function of the hidden layer,

l_i – input variable to the i^{th} node of the input layer,

w_{j0} – bias weight for the j^{th} hidden node.

The modeling process is iterated, where the output of the g_1 function becomes the new input that is weighed with w_{ji} , w_{k0} , and V_j , the hidden value to j^{th} node of the hidden layer. The activation functions are usually continuous, bounded, and nonlinear transfer functions, such as the sigmoid and hyperbolic tangent functions.

Wavelet analysis

Wavelet analysis utilizes a wavelet function known as a mother wavelet defined as:

$$\psi(t) = \int_{-\infty}^{+\infty} \psi(t) dt = 0 \quad (1)$$

and successive wavelets can be derived as:

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

$$b \in R, a \in R, a \neq 0$$

where:

a – scale or frequency factor,

b – time factor,

R – domain of real numbers.

The time scale wavelet transform of a continuous time signal with a finite energy signal $f(t)$, is defined as:

$$W_f(a,b) = |a|^{-\frac{1}{2}} \int_R f(t) \psi^* \left(\frac{t-b}{a} \right) dt \quad (3)$$

where:

$W_f(a,b)$ – wavelet coefficient,

ψ^* – complex conjugate function.

The wavelet transform decomposes the signal $f(t)$ into different components by searching for correlations between the signal and the wavelet function at different scales of a and locally around the time of b , and forming a contour map known as a scalogram. The effort of generating several wavelet coefficients at every possible scale and time can be reduced by constraining the wavelet dilation (a) and translation (b) parameters and defining the discrete wavelet transformation – DWT (Mallat 1989) as:

$$\psi_{m,n}\left(\frac{t-b}{a}\right) = a_0^{-m/2} \psi^* \left(\frac{t-nb_0a_0^m}{a_0^m}\right) \quad (4)$$

where:

m, n – integers that determine the magnitude of wavelet dilation and translation, a_0 – specified dilation step, $a_0 > 1$ (most commonly $a_0 = 2$),

b_0 – location parameter, $b_0 > 0$ (most commonly $b_0 = 1$). For a discrete time series $f(t)$ occurring at a different time – t , where t is an integer time step and assuming $a_0 = 2$ and $b_0 = 1$, the DWT is simplified to:

$$W_f(m, n) = 2^{-m/2} \sum_{t=0}^{N-1} f(t) \psi^*(2^{-m}i - n) \quad (5)$$

where:

$W_f(m, n)$ – wavelet coefficient for the DWT of scale $a = 2^m$ and location $b = 2^m n$,

$f(t)$ – finite time series ($t = 0, 1, 2, \dots, N-1$), $t_{\max} = N$,

M – integer power of 2 ($N = 2^M$),

n – time translation parameter, $0 < n < 2^{M-m} - 1$,

m – magnitude dilation parameter, $1 < m < M$.

The DWT operates two functions (one a high-pass and the second a low-pass filter), and separates the original time series at these two different scales. The wavelet components which identify the high frequencies and fast events, and which can capture small features of interpretational value in the data, are recognized as details (d). The wavelet component or the residual term, representing the background, low frequency informa-

tion (approximation) of data, and which captures longer trend cycles is known as the approximation (A).

Bootstrap technique

Bootstrap resampling is a computational, data-driven simulation method that generates multiple realizations from one dataset of a distribution or process (Efron 1979). Assuming a population of an unknown probability distribution F , where $t_i = (x_i, y_i)$ is a realization drawn independently and identically distributed (i.i.d.) from F , x_i is a predictor vector with y_i , the corresponding output variable, and n is a random dataset sample drawn from F , this results in a random data set sample denoted as $T_n = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$. The empirical distribution function for T_n is \hat{F} with a mass of $1/n$ for each t_1, t_2, \dots, t_n . Similarly, a set of bootstrap samples such as $T^1, T^2, \dots, T^S, \dots, T^S$ can be produced where S is the total number of bootstrap samples, usually ranging from 50 to 200 samples (Efron 1979). In this study, several bootstrap resamples were generated and used to train several different ANN models, and an ensemble forecast was obtained. For each T^s , an ANN model is developed and trained using all n observations and the ANN output, $f_{ANN}(x_i, w_s/T^s)$ is then evaluated using a set A^s of observation pairs $t_i = (x_i, y_i)$ that were not included to generate bootstrap resamples. The performance of the ANNs in these validation tests is subsequently averaged; this also represents the generalization error for the ANN models relative to T_n . This generalization error is denoted as E_0 , and is estimated as (Twomey and Smith 1998):

$$\hat{E}_0 = \frac{\sum_{s=1}^S \sum_{i \in A_s} \left[y_i - f_{ANN}(x_i, w_s / T^s) \right]^2}{\sum_{s=1}^B (A_s)} \quad (6)$$

where:

$f_{ANN}(x_i, w_s / T^s)$ – output of the ANN developed from the bootstrap sample T^s ,
 x_i – particular input vector,
 w_s – weight vector.

Finally, the BANN estimate $\hat{y}(x)$ of all developed ANNs is given by the average of the S bootstrapped estimates (Jia and Culver 2006):

$$\hat{y}(x) = \frac{1}{S} \sum_{s=1}^S f_{ANN}(x, w_s) \quad (7)$$

and the variance is given by:

$$\sigma^2(x) = \frac{\sum_{s=1}^S \sum_{i \in A_s} [y_i - f_{ANN}(x_i / w_s)]^2}{S - 1} \quad (8)$$

Performance indices

The performance of the developed ANN, BANN, WANN and WBANN models were evaluated using four performance indices, namely: coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and peak percentage deviation (P_{dv}) (for additional information on each performance index please see Tiwari and Adamowski 2013).

STUDY AREA AND DATA

Calgary is among the largest cities in Canada, with a population of approximately 1.1 million people (City of Calgary 2011). In 2010, the total per capita water demand in Calgary was 406 liters per day, while residential use was 257 liters per day. The Government of Alberta announced in 2006 that new water licenses for the Bow River Basin will no longer be granted. This situation, in the face of issues such as population growth and climate change, has led to an increased awareness and need for water conservation, demand reduction measures, and consequently to accurate water demand forecasting. The data that was used in this study consisted of average daily water demand, maximum temperature, and total precipitation from 25/3/2004 to 31/12/2006. For the development of the models, the data was divided into three sets: one for training the models (25/03/2004 to 24/03/2005), one for cross-validation (25/03/2005 to 24/03/2006) to check that the models do not overfit, and one for testing the performance (25/03/2006 to 31/12/2006) of the developed models.

MODEL DEVELOPMENT

Artificial neural network (ANN)

Bougadis et al. (2005), and Adamowski et al. (2012) suggested that maximum temperature, total precipitation and water demand itself were very important input variables for water demand forecasting. Therefore, in this study, these three input parameters were considered for optimum model development.

There is no direct method to select the optimum number of inputs for ANN models. A trial and error procedure was adopted by taking the lagged or delayed information of daily time series data of water demand, maximum temperature and total precipitation for one week lead time water demand forecasts. Following this, the optimal number of hidden neurons was selected by a trial and error procedure for each lead time: the model architecture of an ANN was varied by testing 1–10 hidden neurons, and the number of hidden neurons that produced the lowest generalization error was determined to be the optimal structure.

BANN, WANN and WBANN model development

In addition to ANN models, three other ANN-based models were developed in this study: bootstrap-based ANN (BANN), wavelet-based ANN (WANN), and wavelet–bootstrap–ANN (WBANN) models. An ANN model was initially developed using the significant inputs identified after the data was log-transformed and linearly scaled to the range (0, 1). To train the ANN models, a second-order training method, the Levenberg–Marquardt method, was used to minimize the mean squared error between the forecasted and observed water demand values. All the models were developed with Matlab codes using MATLAB® (v. 7.10.0), apart from the bootstrap resamples, which were generated using an Excel add-in (Bootstrap.xla). To develop BANN forecasts, 100 ANN models were developed for each bootstrap resample dataset, and then all the 100 forecasts were used to get an ensemble of forecasts. The WANN model was developed

by inputting the wavelet sub-time series generated using the discrete wavelet components (DWCs). Out of several wavelet functions (Tiwari and Adamowski 2013), the Daubechies wavelet function db5 with three decomposition levels was found to be the most suitable and was used to decompose the time series data for the development of wavelet-based ANN models. All the components of each variable (water demand, maximum temperature, and total precipitation) were used to develop WANN models since all components play a different role in the original time series. The WBANN models are the combination of 100 WANN models developed using 100 resamples of the wavelet sub-time series dataset.

RESULTS AND DISCUSSION

The structure and the performance of the best ANN, BANN, WANN and WBANN models for one week lead time forecasts is shown in Table 1. The observed and forecasted values are shown in Figure 1. For the development of the ANN models, present and past values of maximum temperature (MaxT), total precipitation (TotP) and total water demand (WatDmand) were considered and after a trial and error process the significant input variables for water demand at a weekly lead time step ($t+7$ day) were identified as $WatDmand(t)$ and $TotP(t)$, with six hidden neurons. This best ANN structure was also used to develop the BANN models. After a trial and error procedure (described earlier), the best WANN model was found to have the following inputs: $A3(t)$, $d1(t)$, $d2(t)$, $d3(t)$ of $WatDmand$; $A3(t)$ and $d3(t)$ of $MaxT$, and $A3(t)$ of $TotP$ with 1, 2 and 3 lag time

TABLE 1. Performance of the best models for the testing dataset using ANN, BANN, WANN and WBANN models for 1 week lead time water demand forecasting

Model	Best model structure		Performance indices			
	Input variables	Hidden neurons – HN	R ²	RMSE (ML/day)	P _{dv} (%)	MAE (ML/day)
ANN	WatDmand(t), TotP(t)	6	0.59	60.05	27.38	44.39
BANN	Same as ANN	6	0.56	58.63	17.88	42.53
WANN	A3(t), d1(t), d2(t), d3(t) of WatDmand(t); A3(t) and d3(t) of MaxT(t) and A3(t) of TotP(t) with 1, 2 and 3 lag time variables	4	0.73	45.59	10.85	34.18
WBANN	Same as WANN	4	0.80	40.00	13.68	29.24

WatDmand(t), TotP(t) and MaxT(t) = urban water demand, total precipitation and maximum temperature at time *t*, respectively. *t* = daily time step.

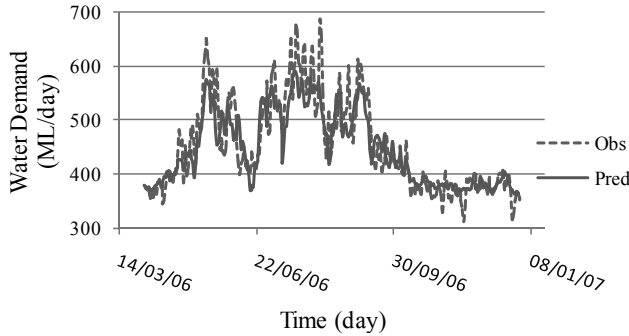


FIGURE 1. Hydrographs of observed and predicted water demand forecasts

variables. The optimum number of hidden neurons was found to be 4. This best WANN structure was also used to develop the WBANN models. The WANN model in terms of R², RMSE, P_{dv} and MAE is significantly better compared (Table 1) to the best ANN and BANN models. The forecast by the WBANN models are the averages of the 100 forecasts by the WANN models trained using different realizations of the training dataset consisting of wavelet sub-time series. The results of the WBANN model in terms of R², RMSE and MAE performance indices for the testing period

is better compared to the ANN, BANN and WANN models. This indicates that, whereas wavelet analysis extracts non-stationarity from the training dataset, bootstrap analysis averages over the error and produces an optimal and accurate result. The WBANN models are trained using 100 realizations of the training dataset and ensemble forecasts are made by averaging 100 forecasts by 100 WANN models trained using 100 realizations of the training dataset. Thus, even if the nature of the training dataset changes, the forecasts will be more reliable and accurate.

SUMMARY AND CONCLUSIONS REFERENCES

Based on four performance indices, it was found that WANN and WBANN models performed considerably better than ANN and BANN models. The better performance of the WANN model for one week lead times compared to the ANN model demonstrates the usefulness of wavelet decomposition. The best ANN and BANN models for one week water demand forecasting considered only water demand and total precipitation to be significant, whereas maximum temperature was not found to be useful in the models. In the case of WANN and WBANN models, after considering wavelet derived sub-time series, all the variable (i.e. water demand, total precipitation and maximum temperature) components were found to be significant and improved the model accuracy considerably. The use of wavelets improved the accuracy of the forecasts, while the use of bootstrapping ensured model robustness along with improved reliability by reducing variance.

This finding is supported by the fact that WBANN models performed well even though they were developed using realizations of very short lengths of training datasets, with these realizations being quite different from each other in their nature. Future research should focus on assessing the effectiveness of WBANN models using longer lengths of datasets (this study focused on situations with limited data availability). The methodology proposed in this study can be useful to make probabilistic forecasts instead of relying on point forecasts.

- ADAMOWSKI J., ADAMOWSKI K., PROKOPHA. 2013: A spectral analysis based methodology to detect climatological influences on daily urban water demand, *Math. Geosci.* 45, 49–68.
- ADAMOWSKI J., CHAN H.F. 2011: A wavelet neural network conjunction model for groundwater level forecasting. *J. Hydrol.* 407, 28–40.
- ADAMOWSKI J., KARAPATAKI C. 2010: Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: evaluation of different ANN learning algorithms. *J. Hydrol. Engg.* 15(10), 729–743.
- ADAMOWSKI J., SUN K. 2010: Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *J. Hydrol.* 390, 85–91.
- ADAMOWSKI J., CHAN H.F., PRASHER S.O., OZGA-ZIELINSKI B., SLIUSARIEVA A. 2012: Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resour. Res.* 48, W01528. doi:10.1029/2010WR009945.
- BOUGADIS J., ADAMOWSKI K., DIDUCH R. 2005: Short-term municipal water demand forecasting. *Hydrol. Processes* 19, 137–148.
- City of Calgary 2011: Civic Census Results. Election and Information Services, Calgary, Canada.
- EFRON B. 1979: Bootstrap methods: another look at the jackknife. *Ann. Statist.* 7(1), 1–26.
- JIA Y., CULVER T.B. 2006: Bootstrapped artificial neural networks for synthetic flow generation with a small data sample. *J. Hydrol.* 331, 580–590.

- MALLAT S.G. 1989: A theory for multi resolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Machine Intel.* 11(7), 674–693.
- TIWARI M.K., ADAMOWSKI J.F. 2013: Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. *Water Resour. Res.* 49 (10), 6486–6507.
- TWOMEY J., SMITH A. 1998: “Bias and variance of validation methods for function approximation neural networks under conditions of sparse data. *IEEE Trans. Syst. Man Cybernet Part C: Appl. Rev.* 28(3), 417–430.

Streszczenie: Tygodniowa prognoza zapotrzebowania na wodę w obszarach miejskich określana metodą hybrydową z wykorzystaniem transformaty falkowej – bootstrap – sztucznej sieci neuronowej. W artykule zaproponowano hybrydowy model (WBANN) wykorzystujący transformatę falkową, bootstrap i sztuczną sieć neuronową do opracowania tygodniowej prognozy zapotrzebowania na wodę w obszarach miejskich przy ograniczonej dostępności danych. Proponowany model WBANN ma na celu poprawę trafności i niezawodności prognozowania zaopatrzenia w wodę. W analizach wykorzystane zostały dobowe wartości maksymalnej temperatury, sumy opadów i zapotrzebowania na wodę z 3-letniego okresu obserwacji. Stwierdzono, że hybrydowy model WBANN jest dokładniejszy od modeli ANN, BANN i WANN i z powodzeniem może być użyty do operacyjnego prognozowania zapotrzebo-

wania na wodę. Model WBANN bardzo skutecznie prognozuje szczytowy popyt na wodę. Dobre wyniki otrzymane z modelu WBANN świadczą o tym, że zastosowana analiza falkowa znacząco poprawiła dokładność modelu, a metoda bootstrapu polepszyła niezawodność (wiarygodność) modelu poprzez prognozowanie ensembledowe. Ocena niepewności z zastosowaniem przedziału ufności wykazała dużą trafność prognoz generowanych przez model WBANN oraz jego przydatność w operacyjnym wykorzystaniu

Słowa kluczowe: sztuczne sieci neuronowe, re-sampling, prognozy ensembledowe, zapotrzebowania na wodę, prognozowanie, transformata falkowa, Kanada

MS. received August 2014

Authors' addresses:

Kaz Adamowski
Department of Civil Engineering, University of
Ottawa, Ottawa
e-mail: Adamowsk@eng.Uottawa.ca

Ousmane Sedou
Department of Civil Engineering, University of
Ottawa, Ottawa
e-mail: oseedou@Ottawa.ca

Jan F. Adamowski
Department of Bioresource Engineering, McGill
University, Quebec, Canada
e-mail: jan.adamowski@mcgill.ca

Bogdan Ozga-Zieliński
Wydział Inżynierii Środowiska
Uniwersytet Warszawski
e-mail: bodgan.ozga-zielinski@is.pw.edu.pl