Journal of Hydroinformatics | 12.2 | 2010

Urban water demand forecasting based on HP filter and fuzzy neural network

Wu Li and Zhou Huicheng

ABSTRACT

Urban water demand is a complex function of socio-economic characteristics, climatic factors and public water policies and strategies. Therefore a combination model is developed based on the multivariate econometric approach which considers these parameters to forecast and manage the urban annual water demand. Firstly, the factors correlative with water demand are selected, then the trend and cyclical components of the factors are calculated by the Hodrick–Prescott (HP) filter method. The multiple linear regression method is applied to simulate the trend components and the fuzzy neural network is built based on the cyclical components, and then the two models are combined to forecast the urban annual water demand. In order to illuminate the model, it is used to forecast the annual water demand of Dalian against actual data records from 1980 to 2007. By comparing with the traditional methods, the preferable model accuracy demonstrates the effectiveness of the fuzzy neural network and multiple linear regression based on the HP filter in forecasting urban annual water demand. After model testing, the sensitivities of the influence factors in the model are analyzed. The results show the model is reliable and feasible, and it also helps to make predictions with less than 10% relative error.

Key words | forecasting, fuzzy neural network, Hodrick–Prescott filter, multiple linear regression, urban water demand

Wu Li (corresponding author)
Zhou Huicheng
School of Civil and Hydraulic Engineering,
Institute of Water Resources and Flood Control,
Dalian University of Technology,
2 Linggong Road,
Ganjinzi District,
Dalian City Liaoning Province 116024,
China
Fax: +86 0411 8470 8517
E-mail: *wuli_water@163.com*

INTRODUCTION

For an increasing number of countries, water scarcity has become a major problem. The even-increasing consumption of water among agricultural, industrial and urban use has led to more vicious competition for water resources, thus impeding social, industrial and rural development of many countries (Hoffmann *et al.* 2006). However, the growing consumption of water is not paralleled by the increasing of water resources, and this fact, in turn, aggravates the competition between regions or countries for water (Ohlsson 1995). Especially in China, more than 400 cities are suffering from insufficient water supply and about 110 cities are facing the more severe situation of water scarcity. The annual urban water shortage amounts to 6 billion cubic meters (Chen *et al.* 2005). Undoubtedly, the water problem is an issue that deserves rigorous management and extreme caution in preventing depletion.

To achieve effective water management in a city, an analysis of water consumption is a must to determine which area needs improvement, how it should be improved and why needs improvement. Meanwhile, the water policies and routines need to be revised to achieve efficiency. According to the time horizon considered, water demand forecasting can be classified as short-, medium- and long-term forecasting. The short-term water demand forecasting is used for real-time water control and allocation, while the medium- and long-term water demand forecasting is used for planning new developments or system expansion, financial planning, capacity planning and so on (Jain *et al.*

doi: 10.2166/hydro.2009.082

2001). This paper aims to establish a long-term forecasting model to forecast the future urban annual water demand. The long-term forecasting is a crucial part in the successful operation of any water distribution system, planning of the projected size of inter-basin water transfer and designing the use of reclaimed sewage.

Historically, water managers have adopted conventional modeling techniques such as regression analysis and time series analysis, or a combination of the two. A lot of work on short-term and medium–long-term water demand forecast modeling using regression, time series analysis and combined methods has been widely reported in the literature.

Regression analysis is the most frequently used statistical technique to model water use from the various related factors such as population size, price of water, average income and annual precipitation. However, this analysis also includes very strict assumptions such as normal distribution and constant variance (Sen *et al.* 2003). Gato *et al.* (2007) presents a new daily demand model incorporating base use values calculated using temperature and rainfall thresholds for East Doncaster, Victoria, Australia was evaluated. The model is based on a postulate that total water use is made up of base use and seasonal use, where base use represents mainly indoor use and is independent of climatic effects such as rainfall and temperature and seasonal use on seasonal, climatic and persistence components. Results revealed the recommended total daily water use model has a combined coefficient determination, R^2 of 80%, which is an improvement on the previous model by Gato *et al.* (2003) with an R^2 of 65%. Babel *et al.* (2007) have proposed multivariate linear, semi-log and log–log approaches for domestic water demand modeling, which considered nine parameters to forecast and manage the domestic water use/demand. The application results have indicated that the level of water pricing, public education level and average annual rainfall were significant variables for domestic water demand.

Time series forecasting, the most widely used approach, relies on the direct identification of patterns existing in historical water demand data. It is assumed that future water use follows the trends in the past and the water use over time is extrapolated into the future by graphical or mathematical means. Thus the change in demand over time may be assumed to follow a linear, logarithmic, exponential or some other function. Jain *et al.* (2001) proposed a relatively new technique of artificial neural networks for short-term water demand forecasting, whose parameters are weekly maximum air temperature, weekly rainfall amount, weekly past water demand, and the occurrence and non-occurrence of rainfall. An example application had shown the artificial neural network models, with an average absolute error of the best artificial neural network model in forecasting of 2.41%, consistently outperformed the regression and time series models, and it also showed the best correlation between the modeled and targeted water demands. Msiza *et al.* (2007) have applied artificial neural networks with multi-layer perceptron and radial basis functions to forecast water demand, in which the input is past water use and annual population size, while the output is monthly and annual water demand. Through comparison, the results have shown that the radial basis functions network is a better model with less validation errors. Ghiassi *et al.* (2008) have presented the development of a dynamic artificial neural network model for comprehensive urban water demand forecasting with a forecasting accuracy of more than 99%. They also examined the effects of including weather information in the forecasting models and found that such inclusion could improve accuracy. But, using time series water demand data, the study has demonstrated that a dynamic artificial neural network model could provide an excellent fit and forecast without reliance upon the explicit inclusion of weather factors.

In recent years, the relative new technique of a combined forecasting model has been proposed as an efficient tool for modeling and forecasting. The model, which takes the advantages of the regression analysis and time series methods, can analyze the relationship between water demand and its exogenous variables such as weather and economic factors, and forecasts the data trend based on the past. Altunkaynak *et al.* (2005) have presented a Takagi–Sugeno (TS) fuzzy method for forecasting future monthly water consumption values from three antecedent water consumption amounts, which were considered as independent variables. After removing the possible trend from the water use time series, the TS fuzzy model was applied to monthly water consumption fluctuations of Istanbul in Turkey. The results have shown that this model predicts water consumption with less than 10% relative error.

At present, research on the combined method are relatively less in number than those on the regression and time series analysis. Even in the study by Altunkaynak *et al.* (2005), the model, which is established in the three antecedent water consumption bases, did not consider the effects of other external variables. Since urban water demand is determined by different factors including the water rights systems, water price, socio-economic development and other natural factors, the urban water consumption is characterized as random, fuzzy and chaotic. Accordingly the fuzzy neural network (FNN) is presented to find a more appropriate relationship between water consumption and its influence factors (Zhou *et al.* 2006). The FNN consists of two parts, where the first part is to calculate the relative membership degree (RMD) for the input data and the second part is a look-up table for mapping input and output patterns. It can solve the uncertainty and fuzzy problem effectively without requiring prior knowledge of the underlying process or making any assumptions about the relationship among variables (Chen & Ji 2005). During the stage of learning, the weight vectors between the hidden and output layers are adjusted by supervised learning to reduce errors. A model with increasing historical data can automatically improve rules, thus enhancing its predictive ability. The FNN has efficient clustering effects from a human-like ability of extracting rules and good simulating results from nonlinear functions. It is a good pattern recognition engine and robust classifier, with the ability to make decisions from fuzzy input data.

According to the trend and fluctuations in the past data, this study mainly develops a long term forecasting model to predict the urban annual water demand. The target of forecasting is urban annual water consumption, and the purpose is to predict water demand amounts for planning level years. Based on the past time series, the model combined multiple linear regression with a fuzzy neural network based on a Hodrick–Prescott filter and is used to model annual water consumption time series, whose inputs are five influence factors including some socio-economic and climate parameters, and the output is the annual water consumption time series. This paper is organized in the following manner. Firstly some socio-economic and climate parameters are selected as model inputs and annual water consumption as the model output. Secondly the model is introduced. It contains three aspects: (1) the Hodrick–Prescott (HP) filter method is introduced, which is used to decompose the time series of annual water demand and its influence factors into trend and cyclical components; (2) the multiple linear regression (MLR) is proposed, which is used to simulate the trend components dataset and (3) the fuzzy neural network (FNN) is presented, which is used to build the model for the cyclical components dataset. Finally the applicability of the model is demonstrated by using an example of annual water consumption in Dalian, China, and incorporating the trend dataset and the cyclical dataset respectively for the MLR and FNN system to obtain the most efficient model configuration.

URBAN WATER DEMANDS

The influence factors of urban annual water demand

The present study is to establish a long-term forecasting model for urban annual water demand based on previous annual water consumption and the time series of its influence factors. Water demands are highly variable and are affected by factors like the size of city, characteristics of the population, the nature and size of commercial and industrial establishments, climatic conditions and cost of supply (Zhou *et al.* 2002), and they have meaningfully increased because of various factors, such as local population growth, global warming, expansion of city greenery coverage, industrial growth and expansion, and consequently a general rise in the living standards. Considering the feasibility of the model, the model inputs are urban annual population, GDP, the annual average temperature, greenery coverage and the previous urban annual water consumption, while the output is the future urban annual water demand, and the factor values are all the annual values. The urban population scale can mainly determine domestic water; GDP has a positive correlativity with urban annual water consumption and can represent the urban economic development level both of industry and agriculture; the urban greenery coverage can represent the urban ecological situation, it can reflect not only the ecological water consumption but also the annual precipitation; the urban annual average temperature also has a large effect on the water consumption: to be specific, the temperature has a positive correlativity with water demand. The higher the temperature is, the more the domestic and irrigation water consumption is; for the water consumption series, it has a certain trend for past several decades, and has great influence on the urban water demand in the future.

Consequently the study selected the following factors with high correlation to forecast the annual water demand, which are urban annual population, urban annual GDP, urban annual average temperature, urban annual greenery coverage and the previous annual water consumption. So the urban annual water demand model can be denoted as

$$W = f(P, G, \mathrm{AT}, A, \mathrm{WU}) \tag{1}$$

where W is the urban annual water consumption time series, f is urban annual water demand function, P is the urban annual population time series, G is the urban annual GDP time series, AT is the urban annual average temperature time series, A is the urban greenery coverage time series and WU is the previous urban annual water consumption time series, $\mathrm{WU}_t = W_{t-1}$, where the subscript t represents the variable for the tth year.

HP filtering for annual water demand and its influence factors

With the passage of time, the annual urban water consumption and its influence factors in a time series are increasing with an obvious trend, but in response to government policy and the influence of some emergent events, those series also show some fluctuations. In order to model the urban annual water consumption and make future predictions, the HP filter technique is adopted to separate the trend component and the cyclical component from the original data. The HP filter is an algorithm for choosing smoothed values for a time series (Hodrick & Prescott 1997), and after filtering, the series is divided into two series, namely a trend component series and a cyclical component series. For annual water demand series and the influence factors series, the HP filter is applied respectively to divide them into two series.

Taking the water demand series W as an example, the trend component series W^{T} can be determined by formula (2):

$$\min\left\{\sum_{t=1}^{N}(W_t - W_t^{\mathrm{T}})^2 + \lambda\sum_{t=2}^{N}[(W_{t+1}^{\mathrm{T}} - W_t^{\mathrm{T}}) - (W_t^{\mathrm{T}} - W_{t-1}^{\mathrm{T}})]^2\right\} \tag{2}$$

where t represents the same as stated above; N represents the sample size of W; the parameter λ is a relative weight of $\sum_{t=1}^{N}(W_t - W_t^{\mathrm{T}})^2$ and $\sum_{t=2}^{N}[(W_{t+1}^{\mathrm{T}} - W_t^{\mathrm{T}}) - (W_t^{\mathrm{T}} - W_{t-1}^{\mathrm{T}})]^2$, $\lambda > 0$. The first term of the equation is the sum of the squared deviations $W^{\mathrm{C}} = W - W^{\mathrm{T}}$ which penalizes the cyclical component. The second term is a multiple λ of the sum of the squares of the trend component's second differences. This second term penalizes variations in the growth rate of the trend component. The larger the value of λ is, the higher the penalty is. Hodrick and Prescott advise that, for quarterly data, a value of $\lambda = 1{,}600$ is reasonable; for yearly data, a value of $\lambda = 100$ is reasonable.

The trend component is subtracted from the actual data and the rest is called the cyclical component, that is

$$W^{\mathrm{C}} = W - W^{\mathrm{T}}$$

Similarly, the influence factors series can also be divided into a trend component series and a cyclical component series by using formula (2). After HP filtering, the water demand influence factors can be denoted as

$$(P, G, \mathrm{AT}, A, \mathrm{WU}) = \{(P^{\mathrm{T}}, G^{\mathrm{T}}, \mathrm{AT}^{\mathrm{T}}, A^{\mathrm{T}}, \mathrm{WU}^{\mathrm{T}}), (P^{\mathrm{C}}, G^{\mathrm{C}}, \mathrm{AT}^{\mathrm{C}}, A^{\mathrm{C}}, \mathrm{WU}^{\mathrm{C}})\}.$$

Combined trend components and cyclical components of the annual water demand and its influence factors respectively, the urban water demand can be divided into two parts to forecast:

$$W^{\mathrm{T}} = f^1(P^{\mathrm{T}}, G^{\mathrm{T}}, \mathrm{AT}^{\mathrm{T}}, A^{\mathrm{T}}, \mathrm{WU}^{\mathrm{T}})$$

$$W^{\mathrm{C}} = f^2(P^{\mathrm{C}}, G^{\mathrm{C}}, \mathrm{AT}^{\mathrm{C}}, A^{\mathrm{C}}, \mathrm{WU}^{\mathrm{C}}).$$

Selecting the influence factors trend components as independent variables of the multiple linear regression, and the cyclical components as the inputs variables of the fuzzy neural network, the two types of multiple linear regression and fuzzy neural network models are developed for annual water demand trend and cyclical components, respectively. The sum of the forecasting results of the trend and cyclical components is the annual water demand value. Thus in order to compare the model of multiple linear regression and fuzzy neural network, the trend components and cyclical components are respectively forecast by the two models.

MULTIPLE LINEAR REGRESSION

Regression analysis was used to investigate the linear relationship between the trend component of the annual water demand and the trend component of the socio-economic and climate variables. Multiple linear regression models were used. The aim of multiple regression analysis is to obtain a linear equation that allows the dependent variable W^{T} to be estimated when the values of the predictive variables $P^{\mathrm{T}}, G^{\mathrm{T}}, \mathrm{AT}^{\mathrm{T}}, A^{\mathrm{T}}, \mathrm{WU}^{\mathrm{T}}$ are known:

$$W^{\mathrm{T}} = b_0 + b_1 P^{\mathrm{T}} + b_2 G^{\mathrm{T}} + b_3 \mathrm{AT}^{\mathrm{T}} + b_4 A^{\mathrm{T}} + b_5 \mathrm{WU}^{\mathrm{T}} \qquad (3)$$

where the parameters $b_0, b_1, \ldots, b_5$ represent the contributions of each independent variable ($P^{\mathrm{T}}, G^{\mathrm{T}}, \mathrm{AT}^{\mathrm{T}}, A^{\mathrm{T}}, \mathrm{WU}^{\mathrm{T}}$) to the estimation of the dependent variable W^{T}.

FUZZY NEURAL NETWORK

The architecture of fuzzy neural network

The fuzzy neural network (FNN), which incorporates fuzzy recognition and a neural network, has both learning and reasoning abilities. Combining the self-learning ability of a neural network, and analyzing the law and fuzzy knowledge in the data, the FNN is used to forecast through learning the data mapping relationship. The FNN model is comprised of user-defined inputs (population, rainfall, temperature, etc.) and desired outputs (annual water demand) that are connected by a set of highly interconnected nodes arranged in a series of layers. These nodes are connected to the user-defined inputs and to the desired output. Figure 1 illustrates the FNN network architecture used in this paper with one input layer with m nodes, one hidden layer with l nodes and one output layer with one node.

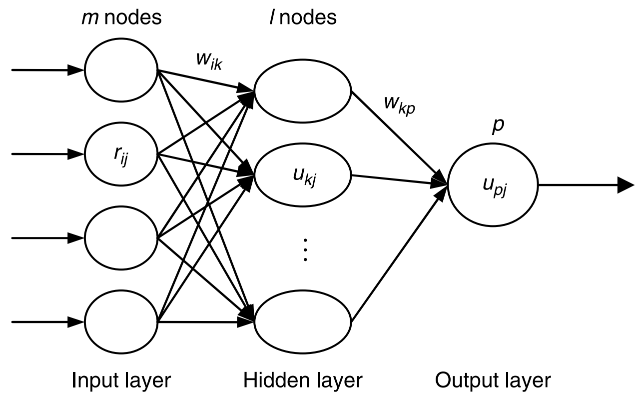

Figure 1 | FNN network architecture.

Fuzzy neural network for water demand forecasting

According to the analysis of annual water demand influence factors, the actual value of the annual water demand is $W = \{W_1, W_2, \ldots, W_N\}$, the feature value for forecast factors is $X = (x_{ij})$, x_{ij} is the feature value of forecast factor i to sample j, $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, N$, m is the node number for the input layer: it equals the number of the factors and N is the sample size of W. In order to classify, the relative membership degree (RMD) of the forecast factors needs to be determined. Here the extensive Gauss function is selected to calculate the RMD:

$$r_{ij} = \exp\left[-\left(\frac{x_{ij} - c_i}{\sigma_i}\right)^2\right] \qquad (4)$$

where c_i and σ_i are respectively the central point and width of forecast factor i:

$$c_i = \bar{x} = \frac{1}{N}\sum_{j=1}^{N} x_{ij} \qquad \sigma_i = \sqrt{\frac{\sum_{j=1}^{N} (x - \bar{x})^2)}{N}}$$

For the FNN, the network input is r_{ij}, which information is translated from the input layer node i to the hidden node, so the input and output is equal in node i, that is $u_{ij} = r_{ij}$; for node k in the hidden layer, the input is $I_{ij} = \sum_{i=1}^{m} w_{ik} r_{ij}$. The fuzzy membership degree (FMD) function is adopted as the activation function between nodes k and p. The output of node k is

$$u_{kj} = \frac{1}{1 + \left[\left[\sum_{i=1}^{m} w_{ik} r_{ij}\right]^{-1} - 1\right]^2} = \frac{1}{1 + (I_{kj}^{-1} - 1)^2} \qquad (5)$$

where w_{ik} is the connection weight for nodes i and k and it satisfies $\sum_{i=1}^{m} w_{ik} = 1$, $w_{ik} \geq 0$.

There is only one node in the output layer, the input is

$$I_{pj} = \sum_{k=1}^{l} w_{kp} u_{kj} \tag{6}$$

where l is the node number for the hidden layer, p is the node number for the output layer, w_{kp} is the connection weight for the hidden layer and output layer, $\sum_{k=1}^{l} w_{kp} = 1$, $w_{kp} \geq 0$. The output is

$$u_{pj} = \frac{1}{1 + \left[\left[\sum_{k=1}^{l} w_{kp} u_{kj}\right]^{-1} - 1\right]^2} = \frac{1}{1 + (I_{kpj}^{-1} - 1)^2} \tag{7}$$

The output of the network is the response for the FNN input r_{ij}. Supposing the expected output of annual water demand j is $M(u_{pj})$, the square error is as follows:

$$E_j = \frac{1}{2}[u_{pj} - M(u_{pj})]^2 \tag{8}$$

When the square error is less than the set value, the annual water demand can be calculated by the inverse function of formula (4):

$$W_{pj} = \sigma_i(-\ln u_{pj})^{1/2} + c_i \tag{9}$$

Principle of FNN training arithmetic

After HP filtering, the trend components and cyclical components of water demand and its influence factors are obtained. The RMD $\{r_{ij}\}$ of each cyclical component for the influence factor calculated by formula (4) is the input of the FNN, and these inputs are converted into outputs by some operation in the FNN. The model performance is as follows:

(1) Firstly, to determine the input layer nodes m and output layer nodes p by data records of influence factors and annual water demand, we endow the parameters w_{ik}, w_{kp} and η in the FNN (see Figure 1) with random initial values and give the maximum calculating time λ and constant ε.

(2) Input $\{r_{ij}\}$ into the FNN and calculate the output u_{pj} by using formulae (5)–(7).

(3) Calculate the network error by formula (8) and adjust the weights to make the E_j minimum. Using the gradient descent algorithm, the weight adjustment is as follows:

$$\Delta_j w_{kp} = -\eta \frac{\partial E_j}{\partial w_{kp}} \tag{10}$$

$$\Delta_j w_{ik} = -\eta \frac{\partial E_j}{\partial w_{ik}} \tag{11}$$

Here, η is the learning coefficient.

According to formula (10),

$$\Delta_j w_{kp} = -\eta \frac{\partial E_j}{\partial I_{pj}} \cdot \frac{\partial I_{pj}}{\partial w_{kp}} \quad \text{where} \quad \frac{\partial I_{pj}}{\partial w_{kp}} = u_{kj}$$

$$\text{Let } \delta_{pj} = \eta \frac{\partial E_j}{\partial I_{pj}} = -\frac{\partial E_j}{\partial u_{pj}} \frac{\partial u_{pj}}{\partial I_{pj}}$$

Based on formula (8), we get

$$\frac{\partial E_j}{\partial u_{pj}} = u_{pj} - M(u_{pj}) \tag{12}$$

$$\frac{\partial u_{pj}}{\partial I_{pj}} = 2u_{pj}^2 \left[\frac{1 - \sum_{k=1}^{l} w_{kp} u_{kj}}{\left[\sum_{k=1}^{l} w_{kp} u_{kj}\right]^3}\right] \tag{13}$$

Substituting the formulae (12) and (13) for δ_{pj}, we get

$$\delta_{pj} = 2u_{pj}^2 \left[\frac{1 - \sum_{k=1}^{l} w_{kp} u_{kj}}{\left[\sum_{k=1}^{l} w_{kp} u_{kj}\right]^3}\right] (M(u_{pj}) - u_{pj}) \tag{14}$$

So the weight adjustment for the hidden layer node k to the output layer node p is

$$\Delta_j w_{kp} = -\eta u_{pj}^2 u_{kj} \left[\frac{1 - \sum_{k=1}^{l} w_{kp} u_{kj}}{\left[\sum_{k=1}^{l} w_{kp} u_{kj}\right]^3}\right] (M(u_{pj}) - u_{pj}). \tag{15}$$

According to formula (11), we get

$$\Delta_j w_{ik} = -\eta \frac{\partial E_j}{\partial I_{kj}} \cdot \frac{\partial I_{kj}}{\partial w_{ik}} = -\eta \frac{\partial E_j}{\partial I_{kj}} r_{ij} \tag{16}$$

Let

$$\delta_{kj}=-\frac{\partial E_j}{\partial I_{kj}}=-\frac{\partial E_j}{\partial u_{kj}}\cdot\frac{\partial u_{kj}}{\partial I_{kj}}=-\frac{\partial E_j}{\partial u_{kj}}2u_{kj}^2\left[\frac{1-\sum_{i=1}^{m}w_{ik}r_{ij}}{\left[\sum_{i=1}^{m}w_{ik}r_{ij}\right]^3}\right],$$

$$\frac{\partial E_j}{\partial u_{kj}}=\frac{\partial E_j}{\partial I_{pj}}\cdot\frac{\partial I_{pj}}{\partial u_{kj}}=-\delta_{pj}w_{kp}$$

so

$$\delta_{kj}=2\delta_{pj}w_{kp}u_{kj}^2\left[\frac{1-\sum_{i=1}^{m}w_{ik}r_{ij}}{\left[\sum_{i=1}^{m}w_{ik}r_{ij}\right]^3}\right].$$

Then the weight adjustment for input layer node i to hidden layer node k is

$$\Delta_j w_{ik}=2\eta r_{ij}w_{kp}u_{kj}^2\left[\frac{1-\sum_{i=1}^{m}w_{ik}r_{ij}}{\left[\sum_{i=1}^{m}w_{ik}r_{ij}\right]^3}\right]\delta_{pj} \tag{17}$$

(4) Return to step (2) until values of the error energy function for the network (formula (8)) is less than the set value ε or the calculating time is bigger than the maximum calculating time N.

APPLICATION

Dalian is located on the east coast of the Euro-Asia continent, the most southern tip of the Liaodong Peninsular in Northeast China. It is an important city of port, trade, industry and tourism. As one of the most heavily developed industrial areas of China, Dalian City is a serious water shortage district, whose total amount of water resources is $32.83 \times 10^8\,\mathrm{m}^3$ and per capita water resource is $575\,\mathrm{m}^3$, only quarter of the per capita water resources in China. In Dalian, groundwater overdraft is serious, and the exploitation and utilization ratio of surface water has reached 40%, which is close to the internationally acknowledged boundary of water utilization. With the development of living standardd and social economy, the water crisis in Dalian is increasingly serious. Consequently, to forecast water demand and allocate the water reasonably is an important basis for Dalian's socio-economic development.

Numerical analysis is performed on annual water demand and the influence factors (population, GDP, annual average temperature, greenery coverage and annual water consumption) obtained from the *Dalian Yearbook*. The annual water demand and the influence factors series are from 1980 to 2007. The preliminary step in investigating is to divide the annual water demand series and the influence factors series into trend components and cyclical components by the HP filter. The trend components of annual water demand and its influence factors are modeled by multiple linear regressions, and the cyclical components are analyzed by the fuzzy neural network. The two components of the annual water demand series and its influence factors series are divided into "training" and "testing" datasets. The former set includes data from 1980 to 2000, i.e. 21 years of annual water demand and its influence factors. The latter contains 7 years of data for testing the model. Finally the performances of the multiple linear regression (MLR) and FNN models are analyzed through comparing the known water consumption data from 2001 to 2007 with the predicted values obtained from different models.

HP filter

Firstly, let $\lambda = 100$. The trend components and cyclical components of both annual water demand series and factors series are calculated by the HP filter formula (2). The analysis results are shown in Figures 2–6.

Figures 2–6 show that the annual water demand and its influence factor values for Dalian are increasing

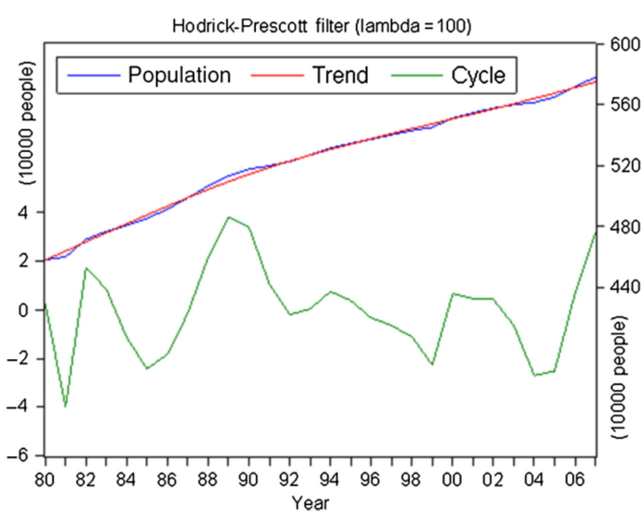

Figure 2 | HP filtered figure of population.

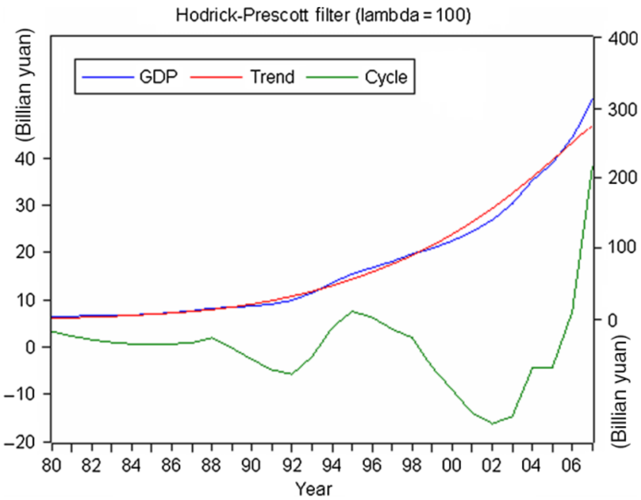

Figure 3 | HP filtered figure of GDP.

with some fluctuations in the past years and, because of government policy and the emergent occurrence, the fluctuations of different factors have different characteristics. Since 1999, Dalian has had drought for five continuous years. And during these years, the government took some water-saving measures to restrict water consumption. Consequently the factor values of GDP, greenery coverage and water consumption are obviously depressed, which can be seen from the cyclical curves in Figures 3, 5 and 6.

After HP filtering, the correlation coefficients between the trend components of annual water demand and factor variables are all increased, but the correlation coefficients

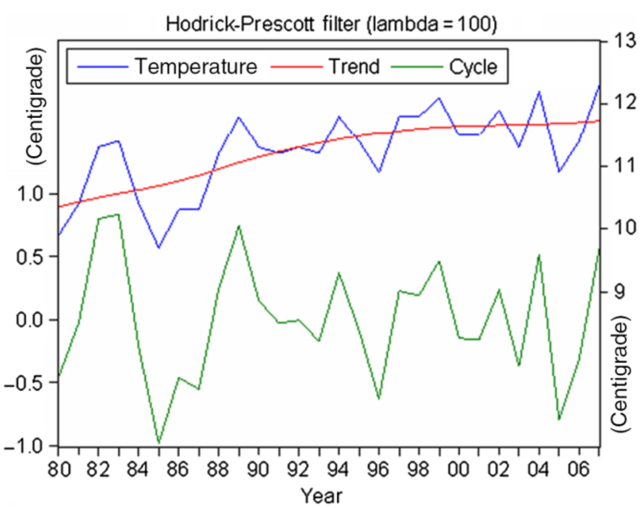

Figure 4 | HP filtered figure of annual average temperature.

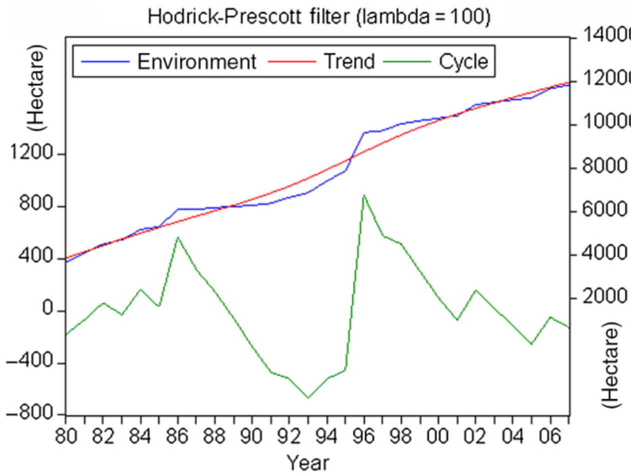

Figure 5 | HP filtered figure of greenery coverage.

between the cyclical components of annual water demand and factor variables are decreased. The correlation coefficient analysis results are listed in Table 1.

Modeling for the two components

After HP filtering, the increasing trend of annual water consumption in the past few years is similar to the trends of influence factors, and the correlations of trend components between annual water demand and each factor are all increased; even the correlation coefficients between annual water demand and population, annual average temperature and greenery coverage are all more than 0.90. The regression equation for trend components, denoted as HP-MLR(T), is established as follows:

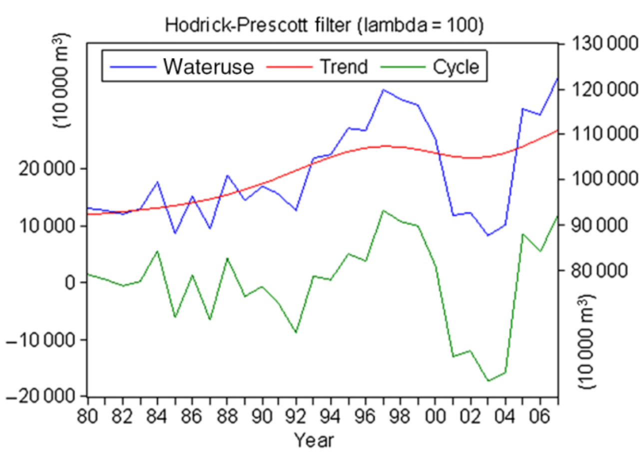

Figure 6 | HP filtered figure of urban annual water demands.

Table 1 | Correlation coefficients between water demand and factors variables for pro- and post-HP filter

Factors	Correlation coefficients with W	The trend components correlation coefficients with W^T	The cyclical components correlation coefficients with W^C
Population	0.524	0.956	0.054
GDP	0.490	0.795	0.611
Annual average temperature	0.431	0.974	0.079
Greenery coverage	0.528	0.936	0.240
Annual water consumption	0.693	0.973	0.508

$$W^{\mathrm{T}} = 101.414P^{\mathrm{T}} + 3.293G^{\mathrm{T}} + 3293.347AT^{\mathrm{T}} - 3.013 A^{\mathrm{T}} + 1.056WU^{\mathrm{T}} - 73790.257 \tag{18}$$

Through analyzing the significance level, the trend components of water demands and its influence factors are especially obvious. For comparison, the FNN model is established to forecast W^{T}, denoted as HP-FNN (T). The actual and simulated values of W^{T} by these two models are shown in Figure 7.

For the trend components with high linear correlation coefficients, both simulated results by HP-MLR(T) and HP-FNN(T) are close to the actual values, and can satisfy the precision requirement.

For the cyclical components, the study adopts the FNN to simulate the cyclical components, in which $(P^{\mathrm{C}}, G^{\mathrm{C}}, R^{\mathrm{C}}, T^{\mathrm{C}}, A^{\mathrm{C}})$ is the input to forecast W^{C}, denoted as HP-FNN (C). Firstly, using formula (4), the RMD of each factor is calculated as the FNN input and the RMD of W^{C} is determined as the FNN expected output $M(W^{\mathrm{C}})$. Then, according to the FNN training arithmetic, the cyclical components of annual water demands are simulated by the cyclical components of the influence factors. Because of too many data, here we only list the RMD of P^{C} and W^{C}:

$$R(P^{\mathrm{C}}) = \begin{matrix} [0.985 & 0.009 & 0.417 & 0.802 & 0.682 & 0.179 & 0.376 \\ 0.995 & 0.251 & 0.013 & 0.033 & 0.718 & 0.989 & 0.999 \\ 0.845 & 0.958 & 0.969 & 0.887 & 0.698 & 0.222 & 0.878] \end{matrix}$$

$$M(W^{\mathrm{C}}) = \begin{matrix} [0.967 & 0.994 & 0.996 & 0.999 & 0.604 & 0.549 & 0.972 \\ 0.513 & 0.742 & 0.914 & 0.994 & 0.807 & 0.295 & 0.978 \\ 0.995 & 0.662 & 0.792 & 0.076 & 0.159 & 0.203 & 0.874] \end{matrix}$$

According to the network input, the input layer has 5 nodes, that is to say, $m = 5$; and let the hidden layer nodes $l = 6$, endowing parameters w_{ik}, w_{kp} and η in the FNN with a random initial value, and giving the maximum calculating time $\lambda = 10{,}000$ and calculated precision $\varepsilon = 10^{-4}$, the network input is $R(P^{\mathrm{C}}, G^{\mathrm{C}}, R^{\mathrm{C}}, T^{\mathrm{C}}, A^{\mathrm{C}})$ and the expected output is $M(W^{\mathrm{C}})$. After determining the network architecture, we can adjust the network parameters and train the network by the model performance steps.

In order to compare the forecasting results, MLR is established for the cyclical components W^{C} and $P^{\mathrm{C}}, G^{\mathrm{C}}$, $\mathrm{AT}^{\mathrm{C}}, A^{\mathrm{C}}, \mathrm{WU}^{\mathrm{C}}$. The regression equation for the cyclical components W^{C}, denoted as HP-MLR(C), is as follows:

$$W^{\mathrm{C}} = 105.075 - 539.58P^{\mathrm{C}} + 43.317G^{\mathrm{C}} + 623.690AT^{\mathrm{C}} + 2.362A^{\mathrm{C}} + 0.312WU^{\mathrm{C}} \tag{19}$$

From Table 1 it is known that the linear correlation is poor between W^{C} and the cyclical components. The minimum correlation coefficient is 0.054. The analyzing results of the regression coefficient significance level have

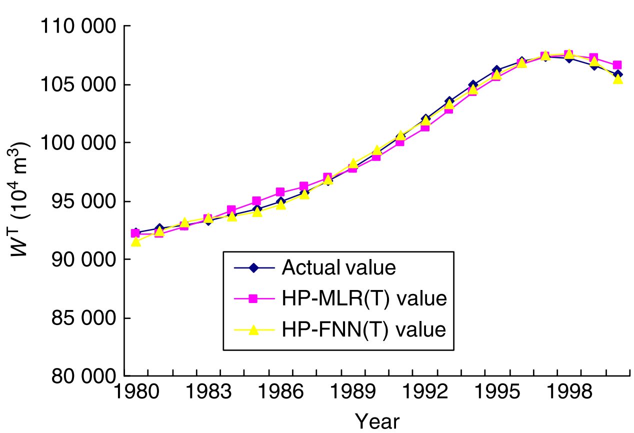

Figure 7 | The actual and simulated cyclical components W^{T} by HP-MLR(T) and HP-FNN(T).

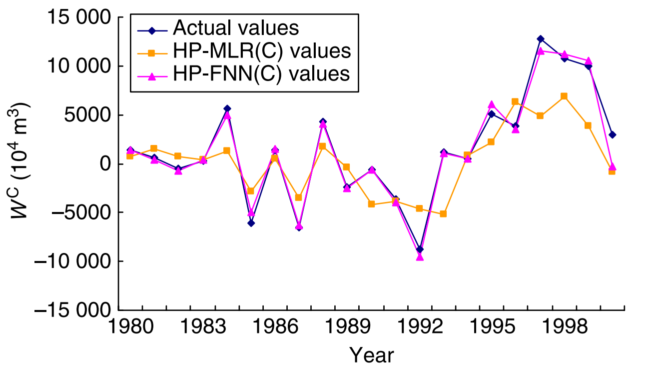

Figure 8 | The actual and simulated cyclical components W^C by HP-MLR(C) and HP-FNN(C).

shown, except for a factor AT^C, the other factors' influences on the annual water demand are all not significant; and through verifying the regression equation significance, we can learn that the regression equation is not significant too. So there exists nonlinear correlativity between the cyclical components of the annual water demand and its influence factors. The results are contrasted with those of HP-FNN(C) and HP-MLR(C) in Figure 8.

It is as expected that HP-FNN(C) gives better simulation results than HP-MLR(C) for nonlinear functions.

Adding the annual water demand trend components to the cyclical components, the urban annual water demand in Dalian is obtained. For comparison, five models are established to simulate annual water demand, which include traditional FNN and MLR and the new models HP-MLR(T) + MLR(C), HP-FNN(T) + FNN(C)

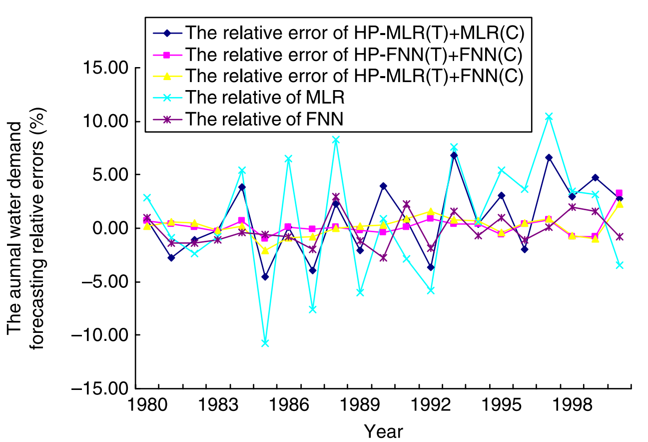

Figure 9 | The annual water demand forecasting relative errors for different forecasting models.

and HP-MLR(T) + FNN(C). The simulation relative errors (RE) of those five models are shown together in Figure 9.

For the urban annual water demand, the simulation precisions of those models based on the HP filter are all increased. The average simulating errors are decreased as follows: from 6.96% of MLR to 2.83% of HP-MLR(T) + MLR(C); from 3.62% of FNN to 1.35% of HP-FNN(T) + FNN(C) and further to 0.59% of HP-MLR(T) + FNN(C). The results of the contrast analysis show that the MLR model is better in prediction for the high linear correlation factors, but the FNN model performs better for both linear and nonlinear correlations between urban annual water demand and its influence factors. And, after HP filtering, the simulation results can be improved and the method proposed in this paper is reasonable.

Model testing

In order to test the presented model HP-MLR(T) + FNN(C), the annual water consumptions from 2001 to 2007 are forecast by the HP-MLR(T) + FNN(C) model based on the corresponding influence factors, and the forecast results are compared with the actual values and the prediction results of the other methods HP-MLR(T) + MLR(C), HP-FNN(T) + FNN(C) and FNN. The actual and forecast values are listed in Table 2.

The maximum relative errors of the forecasting models HP-MLR(T) + FNN(C), HP-MLR(T) + MLR(C), HP-FNN(T) + FNN(C) and FNN are 6.91%, 18.29%, 4.65% and 7.02%. respectively; the average errors are 2.12%, 8.27%, 2.68% and 4.07%, respectively. Comparing the forecast results of the different models in Figure 9 and Table 2, the results show that the simulation results of HP-FNN(T) + FNN(C) are almost as good as for HP-MLR(T) + FNN(C), and both of them are better than the results from the other models, but the forecast results of HP-FNN(T) + FNN(C) are inferior to HP-MLR(T) + FNN(C). Comparing the modeling process of the MLR and FNN, the FNN model is more complex and needs more memory than the MLR method. Therefore the proposed model HP-MLR(T) + FNN(C) is simpler and more appropriate to forecast urban water demand than the others.

Table 2 | The annual water demand values forecasted by four models for Dalian in 2001–2007

Year	Actual value (10^4 m^3)	Forecasting value of HP-MLR(T) + FNN(C) (10^4 m^3)	RE (%)	Forecasting value of HP-MLR(T) + MLR(C) (10^4 m^3)	RE (%)	Forecasting value of HP-FNN(T) + FNN(C) (10^4 m^3)	RE (%)	Forecasting value of FNN (10^4 m^3)	RE (%)
2001	92,127	92,501	0.41	93,245	1.21	93,056	1.01	94,052	2.09
2002	92,772	92,885	0.12	95,548	2.99	94,036	1.36	92,054	0.77
2003	87,705	90,123	2.76	92,914	5.94	91,785	4.65	93,126	6.18
2004	90,027	92,415	2.65	106,495	18.29	93,152	3.47	92,785	3.06
2005	115,800	107,800	6.91	100,221	13.45	111,023	4.13	111,023	4.13
2006	114,500	114,295	0.18	115,198	0.61	113,420	0.94	120,519	5.26
2007	122,600	120,356	1.83	141,504	15.42	126,530	3.21	131,205	7.02

The influence factors' sensitivity analysis for HP-MLR(T) + FNN(C) model

In order to analyze the reasonability of the HP-MLR(T) + FNN(C) model, the sensitivities of the influence factors are analyzed. When the factors of population, GDP, annual average temperature, greenery coverage and the previous water consumption are increased by 5%, the change of annual water demand forecasting values are 2.63%, 0.11%, 1.83%, −1.19%, and 5.25%, respectively. From the results of the sensitivity analysis, we can see the previous annual water consumption has the greatest effect on the urban annual water demand, the population is secondary, and the other factors have a small influence on the annual water demand. Consequently, the parameters of population, GDP, annual average temperature and greenery coverage in the model can be determined by urban development planning or forecasting the varying trend; and the previous annual water demand is an antecedent factor. The model in this study can be used for forecasting the urban annual water demand in a given planning year.

In order to analyze the reasonability of selecting factors, we gradually get rid of the factors with little sensitivity, and take (P, AT, A, WU), (P, AT, WU) and (P, WU) as the model inputs to forecast the annual water demand of Dalian. The maximum relative errors of water demand forecasting values of 1980–2007 are 14.27%, 15.35% and 17.41%, respectively, while the average errors are 3.58%, 4.26% and 5.74%, respectively. Through comparing with the models proposed in the paper, the model with inputs (*P*, *G*, AT, *A*, WU) performs best and it proves that the factors in the model are selected reasonably and comprehensively.

The planning year annual water demand forecasting for Dalian

When forecasting the annual water demand for a planning year, the difficulty is to determine the values of the influence factors. After the HP filter, the trend values of annual water demand and its influence factors are about 0.894–1.196 times the actual value, which shows that the main components for annual water demand and its influence factors are the trend components. For the factors determined difficultly, the values can be obtained through analyzing the increasing trend and fluctuation pattern according to the HP filter results.

Taking 2010 and 2020 as Dalian's planning years, respectively, the factors of population, GDP and greenery coverage can be obtained according to the urban development planning scheduling, and the annual average temperature can be determined by the increasing trend. According to the Dalian development planning project, in 2010, the increasing rate of population is 0.8%, the population will arrive at 6,105,200, GDP will be 438 billion CNY, the greenery coverage will be 12,821 ha and the previous water consumption can use the annual water demand forecasting value in 2009; the future annual average temperature should be determined based on Figure 4. By analyzing the trend curve of annual average temperature in Figure 4, the trend component of the annual average temperature is 11.8. In Figure 4, the fluctuation range of the cyclical component is [−1, 1] and the cyclical component values are about 8.5% of the trend component values. Based on the trend component and fluctuation range of the cyclical component, the interval value of annual average temperature in 2010 is [10.8, 12.8]. Similarly, the values of influence factors in 2020 can be determined. According to the influence factors values in 2010 and 2020, the annual water demands of Dalian in 2010 and 2020 are respectively varying within the intervals [1,473.56 × 10^6, 1,492.95 × 10^6] m^3 and [1,796.35 × 10^6, 1,815.93 × 10^6] m^3 forecast by the above combined forecasting model HP-MLR(T) + FNN(C), and the average values of annual water demands in 2010 and 2020 are respectively 1,483.75 × $10^6\,m^3$ and 1,806.14 × $10^6\,m^3$. By analyzing the forecasting values, we can see that the change of the cyclical component value of the annual average temperature has a small influence on the future annual water demand, and the analyzing result tallies with the conclusion of the sensitivity analysis. Consequently, as a planning forecast, the cyclical component of the annual average temperature can be ignored when its value is small. The forecast results can support the decision-making of allocating the inter-basin water transfer and developing the seawater desalination and reclaimed sewage.

CONCLUSION

This study selects some socio-economic, climate and other related factors to forecast the urban water demand, and in order to improve the forecasting accuracy, the HP filter method is adopted to separate the historical data of the annual water demand and its influence factors into a trend and a cyclical component. After HP filtering, the HP-MLR and HP-FNN techniques are applied to forecast the urban water demand. From the performance of two models based on the HP filter, we can conclude that: (1) after the HP filter, the correlation coefficients of the trend components between annual water demand and the influence factors are all increased and the HP-MLR(T) model is established with high forecasting accuracy. (2) For the poor correlation cyclical components, the HP-FNN(C) model for annual water demand cyclical components is built, and the results show that the HP-FNN(C) has better simulation results for nonlinear functions. (3) Through combining the trend model and cyclical model, the HP-MLR(T) + FNN(C) model is obtained, which considers the linear and nonlinear relationship, respectively. The HP-MLR(T) + FNN(C) model can not only reflect the development trend of water demand and its influence factors, but also analyze their fluctuation characteristics. The application results show that the model HP-MLR(T) + FNN(C) outperforms the HP-MLR(T) + MLR(C) and HP-FNN(T) + FNN(C) techniques used alone; and it also concludes that the better results are obtained when the HP filter is adopted to separate the water demand and its influence factors into trend components and cyclical components. (4) The long-term forecasting model can provide a reference for the water manager to make decisions such as allocating water resources, planning for urban development scheduling, and designing projects of inter-basin water transfer.

ACKNOWLEDGEMENTS

W. Li sponsored by the National Scientific Support Planning project (2007BAB28B01).

REFERENCES

Altunkaynak, A., Özger, M. & Çakmakci, M. 2005 Water consumption prediction of Istanbul city by using fuzzy logic approach. *Water Res. Manage.* **19** (5), 641–654.

Babel, M. S., Das Gupta, A. & Pradhan, P. 2007 A multivariate econometric approach for domestic water demand modeling: an application to Kathmandu, Nepal. *Water Res. Manage.* **21**, 573–589.

Chen, S. & Ji, H. 2005 Fuzzy optimization neural network approach for ice forecast in the Inner Mongolia reach of the Yellow River. *Hydrol. Sci. J./J. Sci. Hydrologiq.* **50** (2), 319–330.

Chen, Y., Zhang, D., Sun, Y., Liu, X., Wang, N. & Savenije, H. H. G. 2005 Water demand management: a case study of the Heihe River Basin in China. *Phys. Chem. Earth* **30**, 408–419.

Gato, S., Jayasuriya, N. & Hadgraft, R. 2003 A simple time series approach to modeling urban water demand. In *Proceedings of the 28th International Hydrology and Water Resources Symposium, Wollongong, NSW*, pp. 10–14. ACT Institution of Engineers, Barton, Australia.

Gato, S., Jayasuriya, N. & Roberts, P. 2007 Temperature and rainfall thresholds for base use urban water demand modeling. *J. Hydrol.* **337**, 364–376.

Ghiassi, M., Zimbra, D. K. & Saidane, H. 2008 Urban water demand forecasting with a dynamic artificial neural network model. *J. Water Res. Plann. Manage.* **134** (2), 138–146.

Hodrick, R. & Prescott, E. C. 1997 Postwar US business cycles: an empirical investigation. *J. Money Credit Banking* **29** (1), 1–16.

Hoffmann, M., Worthington, A. & Higgs, H. 2006 Urban water demand with fixed volumetric charging in a large municipality: the case of Brisbane, Australia. *Aust. J. Agric. Res. Econ.* **50**, 347–359.

Jain, A., Varshney, A. K. & Joshi, U. C. 2001 Short-term water demand forecast modeling at IIT Kanpur using artificial neural networks. *Water Res. Manage.* **15**, 299–321.

Msiza, I. S., Nelwamondo, F. V. & Marwala, T. 2007 Water Demand Forecasting Using Multi-Layer Perceptron and Radial Basin Functions. In *Proceedings of International Joint Conference on Neural Networks*, pp. 13–18. IEEE, Orlando, FL.

Ohlsson, L. 1995 *Hydropolitics–Conflicts Over Water as a Development Constraint*. Zed Books, University Press, London.

Sen, Z., Altunkaynak, A. & Ozger, M. 2003 Autorun persistence of hydrologic design. *J. Hydrol. Eng.* **8** (6), 329–338.

Zhou, S. L., McMahon, T. A., Walton, A. & Lewis, J. 2002 Forecasting operational demand for an urban water supply zone. *J. Hydrol.* **259**, 189–202.

Zhou, H., Wu, L. & Guo, Y. 2006 Mid- and long term hydrologic forecasting for drainage area based on WNN and FRM. In *Proceedings of Sixth International Conference on Intelligent Systems Design and Applications*, Vol. 1, pp. 7–12. IEEE, Jinan, China.

First received 14 November 2008; accepted in revised form 3 March 2009. Available online 21 November 2009