Application of short-term water demand prediction model to Seoul

C.N. Joo, J.Y. Koo and M.J. Yu

Department of Environmental Engineering, University of Seoul, 90 Jeonnongdong Dongdaemungu, Seoul, 130-743, Korea

Abstract To predict daily water demand for Seoul, Korea, the artificial neural network (ANN) was used. For the cross correlation, the factors affecting water demand such as maximum temperature, humidity, and wind speed as natural factors, holidays as a social factor and daily demand 1 day before were used. From the results of learning using various hidden layers and units in order to establish the structure of optimal ANN, the case of 3 hidden layers and numbers of unit with the same number of input factors showed the best result and, therefore, it was applied to seasonal water demand prediction. The performance of ANN was compared with a multiple regression method. We discuss the representation ability of the model building process and the applicability of the ANN approach for the daily water demand prediction. ANN provided reasonable results for time series prediction.

Keywords Artificial neural network; cross correlation; multiple regression; short-term water demand

Water Science and Technology **Vol 46 No 6–7 pp 255–261**

Introduction

Water supply in a metropolis like Seoul requires integrated control of the components of the water system. For the effective management of complicated systems, characteristics of water demand and uncertainties in natural phenomena and social activities should be considered.

Water demand modeling plays a key role in urban water resources planning and management, for example, water and wastewater facility planning and scheduling, optimal operation of existing systems, evaluation of water conservation programs, and assessment of water pricing policy. Since the residential water use study by Howe and Linaweaver (1967), a number of water demand models have been proposed in the context of linear multiple regression (e.g. Morgan and Smolen, 1976; Hansen and Narayanan, 1981) and time series analysis (e.g. Maidment and Parzen, 1984). Water use, together with socioeconomic and climatic variables which have potential influences on water use, have been collected and examined. Socioeconomic variables such as population, income, water price, and housing characteristics are postulated to impose long-term changes on water use patterns; while climatic variables such as precipitation and temperature induce short-term seasonal variations.

The purposes of this study are to develop a prediction model for short-term water demand with consideration of the relationship between regional characteristics and water demand, and to accomplish efficient distribution management for quantitatively-stable, qualitatively-safe water supply, as a part of water operation planning in a city.

Methods

Study area

The target city for this study is Seoul, in which the waterworks have originated from the Tukdo water treatment plant (water treatment capacity, 12,000 m^3/day) constructed in 1908 for supplying 100 lpcd for 125,000 persons. At present, through several expansions of facilities, the current waterworks system of Seoul can supply 6.8 million m^3/day with

10 water treatment plants and 23 reservoirs having a volume of more than 10,000 m^3. The water sources of Seoul are along the Han River, mainly downstream of the Paldang dam.

Seoul is a densely-populated city with a population of ten million. Seoul is located in the temperate zone, which is characterized by four distinct seasons. The average annual temperature of Seoul is 12.9°C. However, it has a wide annual temperature range, with the highest temperature in summer reaching 36.1°C, and the lowest temperature in winter falling to –13.7°C. The annual precipitation in Seoul is 1,210.2 mm. About 70 percent of the annual precipitation falls from June to September. The public water supply systems in Seoul serve 99.9% of the total population and the average water supply per capita per day is 444 L. Figure 1 shows the locations of water treatment plants and reservoirs in Seoul.

Artificial neural network (ANN)

ANN is a computational method inspired by studies on the brain and nerve systems (Thomas and James, 1998). Typically, ANN consists of a set of layered processing units and weighted interconnections (Figure 2). Learning in ANN involves adjusting the weights of interconnections (Rumelhart and McClelland, 1986; Rumelhart *et al.*, 1994).

The ANN approach is winning popularity in the analysis of water resource phenomena. In principle, ANN can compute arbitrary precision values that can be represented as a mapping between vector spaces. In practice, it is especially useful for mapping problems that are tolerant of high error rates and have abundant data available but to which hard and fast rules cannot be easily applied.

ANN can be visualized as a set of interconnected nodes arranged in the layers. The input layer contains one node for each of the input variables. A numeric weight is associated with each of the internode connections. The nodes themselves process the values entering the node to produce an output value. Learning in neural networks consists of presenting the network with known input and output values and adjusting the weight between the nodes so that a reasonable approximation of the known output value is calculated. The process is repeated on each case in the learning set until the output of the network agrees with the known output within a specified accuracy range (Figure 3).

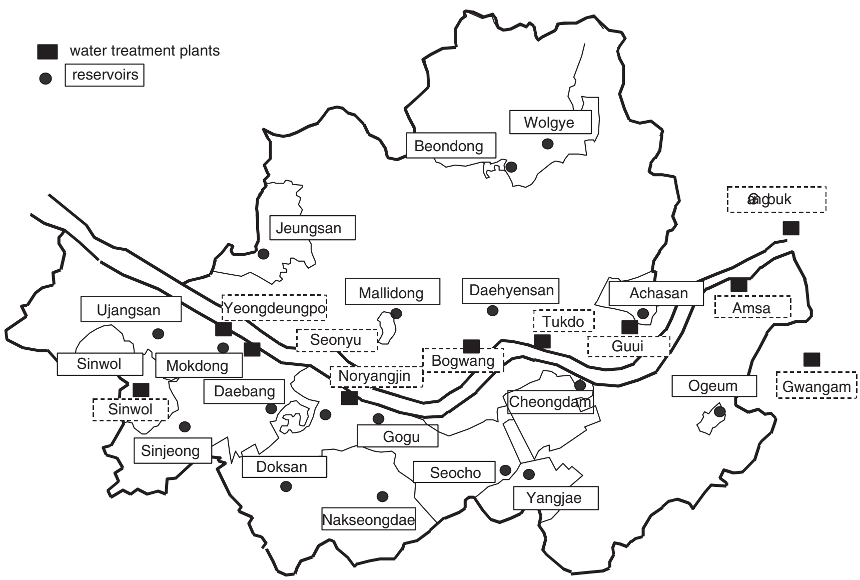


Figure 1 The locations of water treatment plants and reservoirs in Seoul

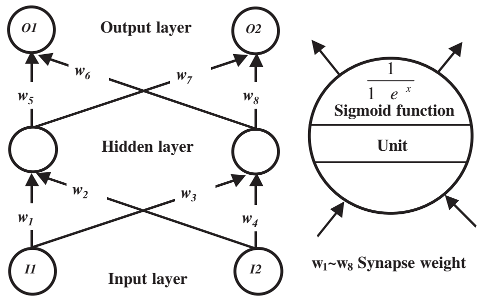


Figure 2 A three-layered artificial neural network and transfer function in unit

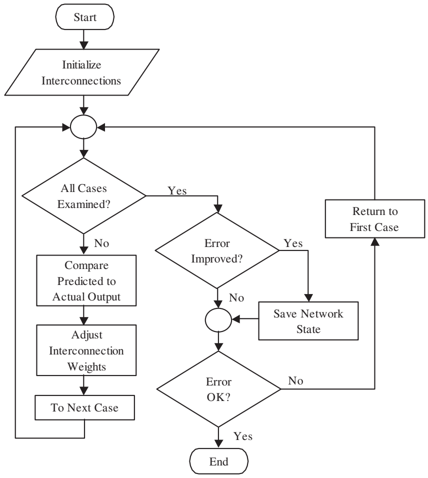


Figure 3 Neural network training procedure

The performance of ANN is compared with a multiple regression method. Multiple regression is an equation for estimating a dependent variable Y with the independent variables X_1, X_2, and so on.

Results and discussion

Analysis of affecting factors

Seasonal indices of Seoul and each water supply district were made and applied to the study for estimating fluctuation of seasonal water demand. From the results, it was possible that one year was considered as 4 seasons; January–March (Winter), April–June (Spring), July–September (Summer), and October–December (Fall).

A cross correlation study is a method analyzing two independent time series. If there are two time series, x_t and y_t, a cross correlation coefficient ($r_c(k)$) is as follows (Box and Jenkins, 1976):

$$r_c(k) = \frac{\sum_{t=1}^{n-k}\{(x_t - \overline{x_t})(y_{t+k} - \overline{y_{t+k}})\}}{\sqrt{\sum_{t=1}^{n-k}(x_t - \overline{x_t})^2 \sum_{t=1}^{n-k}(y_{t+k} - \overline{y_{t+k}})^2}} \tag{1}$$

where,
n = number of data, k = time lag

In Eq. (1), $r_c(0)$ is a cross correlation coefficient of data, which was measured at the same time. And $r_c(1)$ is measured one time later between y and x. In Eq. (1), namely, a cross correlation coefficient represents the strength of linear relationship between x_t, and y_{t+k} which is measured at k time later.

Seasonal data and natural phenomena, such as temperature (max., mean, min.), precipitation, humidity, wind speed, and cloud amount, were applied to the study for investigating correlation with water demand. Results are shown in Figure 4.

It was shown that water demand fluctuation in spring and summer was considerably affected by temperature and somewhat affected by wind speed and humidity. Water demand fluctuation in fall was less affected by temperature than in spring and summer, and in winter. it was little affected by temperature.

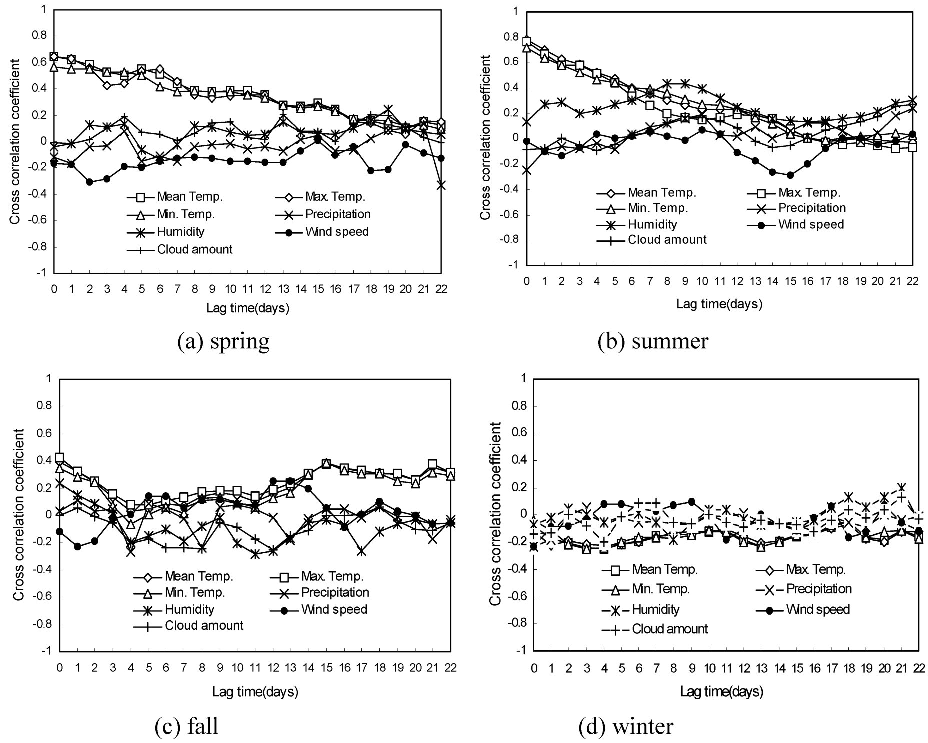


Figure 4 Cross correlation correlogram

On the other hand, from the previous research (Joo, 2001), there was some difference in influence of the factors according to water district. Among the factors, temperature affected all the seasons in Y water district. Wind speed had a greater effect in spring, and water demand and humidity had less influence in winter.

It is known that water demand is also affected by natural phenomena and social factors such as holidays. Selected factors of each seasonal water demand prediction model were as follows (Table 1).

Determination of ANN structure

Input and output layers of artificial neural networks are determined by input and output factors, respectively. But the characteristics of hidden layers are arbitrary and, therefore, to make an optimal model, the number of hidden layers and units must be determined.

In this study, an optimal model was made using water demand in summer, the most fluctuating season, rather than modeling all seasons. The same model structure was then used to consider other seasonal fluctuations.

An artificial neural network uses the data set between 0 and 1 and this data set can be normalized as in Table 2. The water demand and the previous day water demand were divided by water treatment capacity and standardized.

Table 1 Input data

Season	Input data
Spring	Max. Temp., Holiday, Previous water demand, Wind speed
Summer	Max. Temp., Holiday, Previous water demand
Fall	Max. Temp., Holiday, Previous water demand
Winter	Max. Temp., Holiday, Previous water demand, Humidity

The hidden layer structure was studied by using summer seasonal input factors such as maximum temperature, holiday, and the previous day water demand. The number of training runs was 200, and through this proceeding, the correlation coefficient and mean absolute error (MAE) were found. Results are in Table 3.

As the results of learning in various layers and units in order to establish the structure of the optimal ANN, the case of 3 layers and number of unit has same number of input factor showed the best result and, therefore, it was applied to other seasonal water demand prediction.

Estimation of daily water demand

The model which predicts a complicated time series of water demand was considered. Through the comparative analysis of models using daily data, the characteristics of each model were understood. Figure 5 shows the predicted values by ANN compared with those of the multiple regression model. Correlation coefficients and mean absolute errors were calculated for evaluating the fittings of the model, and summarized in Table 4.

As shown in Table 4, the range of mean absolute error for each seasonal water demand prediction was 1.30–1.71%, and this range means that predicted values described measured values very well.

Maximum range of error was 4.89–8.21%, and in the results the correlation coefficients of the multiple regression model and ANN were 0.6–0.86 and 0.61–0.87, respectively. From the values, it can be explained that the multiple regression model and artificial neural network are analogous. Correlation of the model in fall was lower than in other seasons.

Figure 6 and Table 4 show the next month's water demands predicted using the presumed equation which uses 3-month-scale data. The range of mean absolute error was very small (1.55–3.04%), and it is known that the hitting ratio within 5% against the measured value was within the range of 90–100%.

Table 2 Normalization of data

Factor	Normalization
Water demand (y_t)	$y_t = Y_t/600{,}000$
Max. temp. (x_{1t})	$x_{1t} = (X_{1t} + 15)/60$
Holiday (x_{2t})	$x_{2t} = X_{2t}$
Previous water demand (x_{3t})	$x_{3t} = X_{3t}/600{,}000$
Wind speed (x_{4t})	$x_{4t} = X_{4t}/10$
Humidity (x_{5t})	$x_{5t} = X_{5t}/100$

Table 3 Estimation of hidden layer structure

Hidden layer structure	1 layer 3 unit	1 layer 4 unit	1 layer 5 unit	1 layer 6 unit	2 layer 3 unit	3 layer 3 unit
MAE (%)	1.796	1.797	1.810	1.751	1.685	1.685
Correlation coefficient (R)	0.818	0.820	0.817	0.827	0.840	0.842

$$R = \frac{\sum_{t=1}^{n}\{(y_t - m_y)(\hat{y}_t - \hat{m}_y)\}}{\sqrt{\sum_{t=1}^{n}(y_t - m_y)^2 \sum_{t=1}^{N}(\hat{y}_t - \hat{m}_y)^2}} \qquad MAE = \frac{\sum_{t=1}^{n}\left| (y_t - \hat{y}_t)/y_t \right|}{n}$$

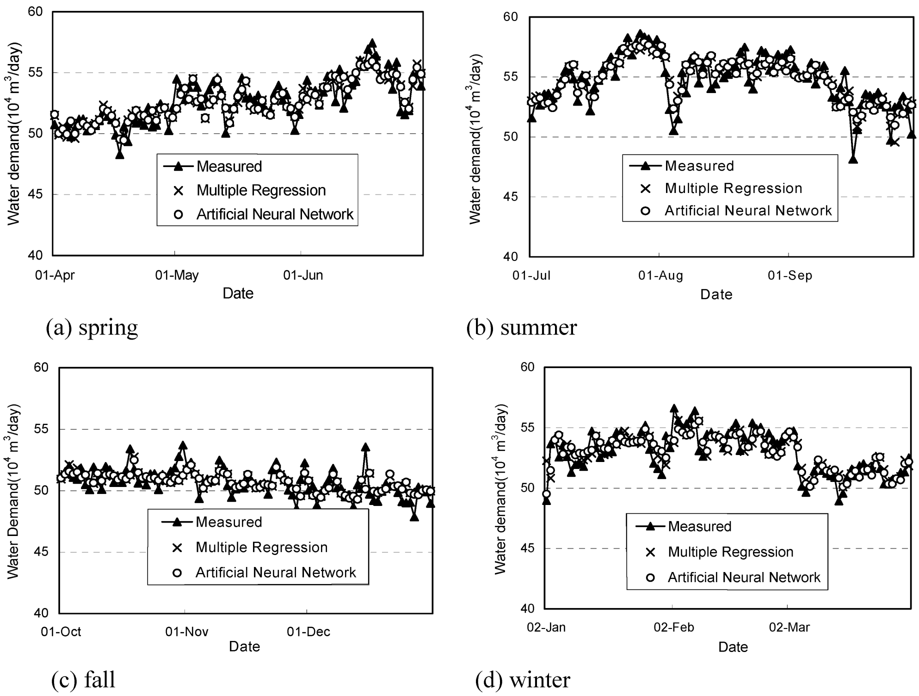


Figure 5 Comparison between the measured and predicted values by the multiple regression and ANN

Table 4 Comparison of estimations by a multiple regression model and an artificial neural network model

Model	Estimation item	Spring	Summer	Fall	Winter
Multiple regression	MAE (%)	1.46	1.71	1.33	1.57
	R	0.86	0.83	0.60	0.78
Artificial neural network	MAE (%)	1.38	1.68	1.30	1.48
	R	0.87	0.84	0.61	0.82

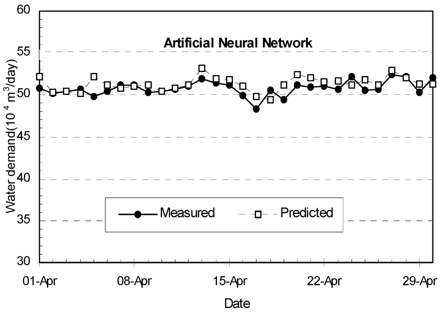
Conclusions

Seasonal water demand prediction models were constructed with the artificial neural network model and compared with past measured data for estimating characteristics of time series and prediction accuracy, and suitability of seasonal short-term water demand prediction was considered.

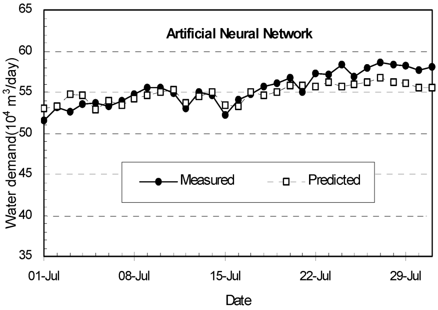
The seasonal water demand prediction models consisted of the seasonal factors such as temperature, humidity, wind speed, holidays and previous daily water demand. Those factors had a high cross correlation with the constructed model.

The models were used to predict the next month's water demands for the comparison of measured values with predicted ones, showing that they could possibly explain the fluctuation of water demand.

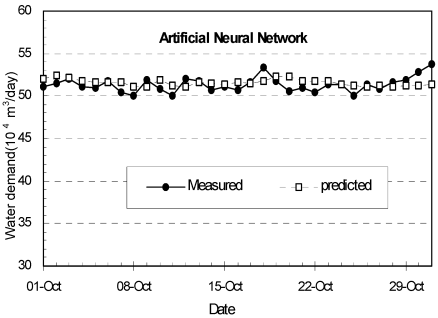
From these results, if accurate water demand, with factors and GIS(Geographic Information System), were estimated, it would be possible to make a more accurate water demand prediction model.



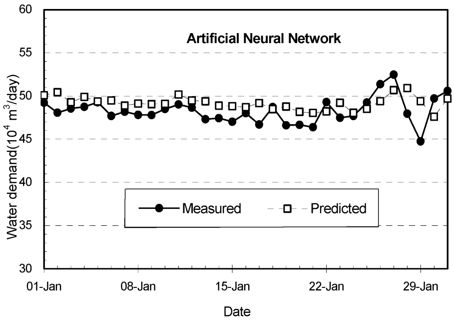
(a) April



(b)July



(c) October



(d) January

Figure 6 Comparison between measured and predicted values by ANN

Table 5 Comparison of predicted values by an artificial neural network

Model	Item	April	July	October	January
Artificial neural network	MAE (%)	1.59	2.02	1.55	3.04
	Hitting Ratio (± 5%)	100%	100%	100%	100%

References

Box, G.E.P. and Jenkins, G.M. (1976). *Time series analysis, forecasting and control*, Holden-Day, San Francisco, Calif.

Hansen, T.D. and Narayanan, R. (1981). A monthly time series model of municipal water demand, *Water Resour. Bull.*, 17(4), 578–585.

Howe, C.W. and Linaweaver, Jr. F.P. (1967). The impact of price on residental water demand and its relation to system design and price structure, *Water Resour. Res.*, 3(1), 13–32.

Joo, C.N. (2001). *Development and application of short-term water demand prediction model for distribution management*, PhD thesis, Department of Environmental Engineering, University of Seoul.

Maidment, D.R. and Parzen, E. (1984). Cascade model of monthly municipal water use, *Water Resour. Res.*, 20(1), 15–23.

Morgan, W.D. and Smolen, J.C. (1976). Climate indicators in the estimation of municipal water demand, *Water Resour. Bull.*, 12(3). 511–517.

Rumelhart, D.E., Widrow, B. and Lehr, M.A. (1994). *The basic ideas in neural networks*, Commun. ACM, 37(3), 87–92.

Rumelhart, D.E. and McClelland, J.L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*, MIT Press Cambridge, Mass., 344.

Thomas, A.C. and James, M.E. (1998). Using neural networks to assess the influence of changing seasonal climates in modifying discharge, dissolved organic carbon, and nitrogen expert in eastern Canadian rivers, *Water Resources Research*, 34(3), 447–455.