Water Resour Manage (2007) 21:399–408
DOI 10.1007/s11269-006-9022-6

Forecasting surface water level fluctuations of lake van by artificial neural networks

Abdüsselam Altunkaynak

Received: 26 April 2005 / Accepted: 16 February 2006

Abstract Lake Van in eastern Turkey has been subject to water level rise during the last decade and, consequently, the low-lying areas along the shore are inundated, giving problems to local administrators, governmental officials, irrigation activities and to people's property. Therefore, forecasting water levels of the Lake has started to attract the attention of the researchers in the country. An attempt has been made to use artificial neural networks (ANN) for modeling the temporal change water levels of Lake Van. A back-propagation algorithm is used for training. The study indicated that neural networks can successfully model the complex relationship between the rainfall and consecutive water levels. Three different cases were considered with the network trained for different arrangements of input nodes, such as current and antecedent lake levels, rainfall amounts. All of the three models yields relatively close results to each other. The neural network model is simpler and more reliable than the conventional methods such as autoregressive (AR), moving average (MA), and autoregressive moving average with exogenous input (ARMAX) models. It is shown that the relative errors for these two different models, are below 10% which is acceptable for engineering studies. In this study, dynamic changes of the lake level are evaluated. In contrast to classical methods, ANNs do not require strict assumptions such as linearity, normality, homoscadacity etc.

Keywords Hydrologic budget · Lake level · Neural networks · Prediction

Introduction

People living near lakes and seas observe water level changes from year to year to be highest in summer, and lowest in winter season. These changes do not show simple periodic fluctuations. Streamflow data is affected by various factors influencing the water budget of a drainage area, and lake-water level fluctuations represent the end result of the complex interplay of

A. Altunkaynak (✉)
Istanbul Technical University, Faculty of Civil Engineering, Hydraulics Division, Maslak 34469, Istanbul, Turkey
e-mail: altunkay@itu.edu.tr

the various water balance components. Among those components are the flow of incoming or outgoing rivers and streams, direct precipitation onto the lake surface and the groundwater exchange. Furthermore, meteorological factors, including precipitation over the lake drainage area, evaporation from the lake surface, wind velocity, humidity and temperature in the adjacent lower atmosphere, play significant roles in lake water level fluctuations.

Since gradual (trend) or abrupt (shifts) climatic change problems have gained particular attention in recent years, most of the researchers on lake level changes are concerned with meteorological variables such as temperature and precipitation (Kadıoğlu *et al.*, 1997; Şen *et al.*, 1999). On a large scale, climatic changes can affect atmospheric circulation patterns, storm frequencies and intensities. All these factors, in turn, have major implications on water availability and quality. Projections of future climate using computer-based models are very uncertain and the results from different models vary. Nevertheless, the best current estimates indicate that average global temperatures will rise by 1 °C–6 °C over the next century and that average global precipitation will increase by as much as 15 percent, with increase in some regions and decrease in others (Frederick and Gleick, 2001). It is widely agreed that temperatures will increase and precipitation patterns will change, but the timing and extent of these future changes still preserve their uncertainty.

This study deals with lake water levels based on Kang *et al.* (1993) who used ANNs and ARIMA models to predict daily and hourly streamflows. Their preliminary study concluded that ANNs are useful tools for forecasting streamflows. ANNs have been used intensively in water resources domain (Jain *et al.*, 2001; Cancelliere *et al.*, 2002; Muleta and Nicklow, 2005; Agarwal and Singh, 2004; Kumar *et al.*, 2004; Şen *et al.*, 2004; Sivapragasam and Muttil, 2005). Huge economic damages occur because of excessive rising or descent of lake water levels. This problem therefore has become an important subject for hydrological studies. It is now possible to determine the water level changes with some certainty by the help of water budget method, which was developed to explain changes in lake's water levels, (Kadıoğlu *et al.*, 1999).

Şen *et al.* (2000) identified suitable models for estimating lake level fluctuations and their parameters including the trend, periodic and stochastic parts. A second order Markov model is found suitable for the stochastic part. Altunkaynak *et al.* (2003) established the triple diagram model of lake levels which is suggested as a replacement of the second order Markov process. Persistence was evaluated by triple diagram model instead of using the first and second order autocorrelation coefficients.

In this study, a three-layered ANN is used to predict Lake Van monthly water level fluctuations from past values. Its results are compared with the classical ARMAX model results.

Time series modeling

Many series, actually encountered in hydrology, exhibit nonstationary behaviour and in particular do not vary about any mean. Water levels of the Van Lake show a nonstationary behaviour. As seen from Figure 1, the mean values change with time.

Box *et al.* (1994) stated that the forecasting of future values of a time series from current and past values is one of the important areas of application in dynamic modeling. Most of the models of levels and flow series of the Great Lakes have assumed stationarity of time series using either Markov or ARIMA models as stated by Slivitzky and Mathier (1993). Multivariate models using monthly lake levels failed to adequately reproduce the statistical properties and persistence of basin supplies (Loucks, 1989; Iruine and Eberthardt, 1992). On

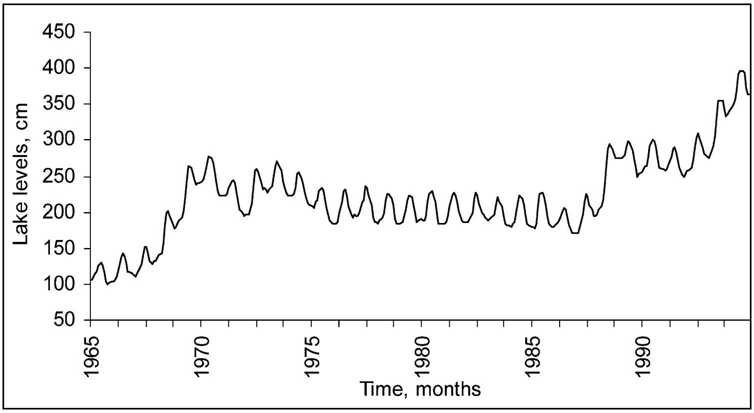

Fig. 1 Lake Van mean monthly surface water level from 1965 to 1994

the other hand, Vaziri (1997) used artificial neural networks (ANN) and ARIMA model to predict water levels of Caspian Sea. He observed that on average the ANN model underestimates the levels by 3 cm whereas the ARIMA model overestimates by 3 cm. He has taken 12 nodes in the input layer which are in a sequence representing the lagged Caspian Sea monthly surface levels from Y_{t-1} to Y_{t-12}.

Artificial neural network (ANN)

An ANN is a massively parellel-distributed information processing system that simulates the working of the neuron network in human brain. Neurons are responsible for the human capacity to learn and this significant property is used in machine learning in artificial neural networks.

Although the concept of artificial neurons was first introduced by McCulloch and Pitts (1943), major applications of ANNs have arisen after the development of the error back propagation method of training by Rumelhart *et al.* (1986). ANN research, following this development, has resulted in successful solution of some complicated problems not easily solvable by traditional methods (Ellis *et al.*, 1993; Suen and Eheart, 2003; Jain and Chalisgaonkar, 2000).

A neural network consists of a number of simple processing elements, called nodes. A simple and general representation of a processing element is shown in Figure 2. The ANNs are composed of network architecture and mathematical functions. The organization of nodes according to a particular arrangement is formed the architecture. The nodes are generally arranged in layers which provide an information flux from input layer to output layer. There can be several hidden layers between input and output layers. The hidden layers increase the network's ability to model more complex events. A three-layer feed-forward ANN along with a typical processing element is shown in Figure 3. The nodes in one layer are connected to those in the next, but not to those in the same layer. Each node computes some function of its input and passes the result to connected units in the networks. Thus the output of a node is related with the inputs and corresponding weights. The strength of the signal passing from one neuron to the other depends on the weight of the interconnection. The number of hidden

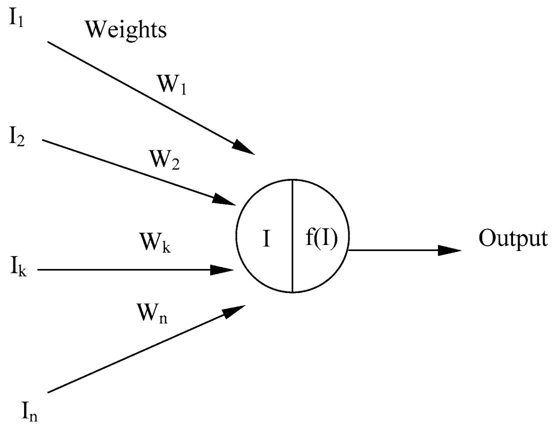

Fig. 2 Simple model of a neuron

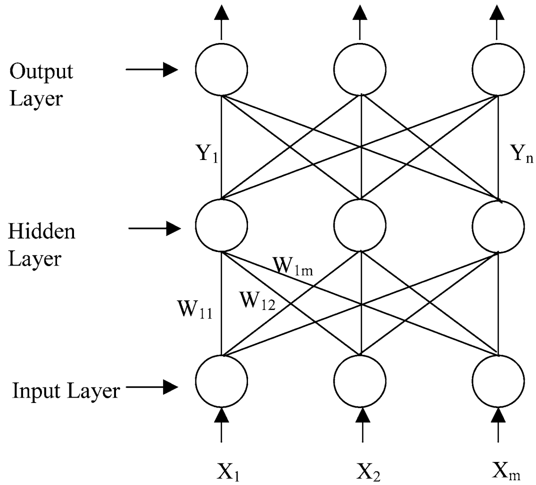

Fig. 3 A multi-layer perceptron

layers and the number of nodes in each hidden layer are usually determined by trial-and-error procedure.

After McCulloch and Pitts (1943), development of artificial neural networks has gained acceleration. Within the last decade, it has experienced a huge resurgence due to the development of more sophisticated training algorithms. The ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000 a,b) has come out with a set of this papers which investigate the role of ANNs in hydrology. The first part is an introduction to ANNs and offers a brief comparison of the nature of ANNs and other modeling philosophies in hydrology. The second part of the series deals with applications of ANNs. It is demonstrated that ANNs are robust tools for modelling many nonlinear hydrologic processes such as rainfall-runoff, streamflow, ground-water management, water quality simulation, and precipitation.

An ANN does not have any knowledge at the beginning. Learning process starts on entering data into the input layer of the network. The error back-propagation (BP) algorithm is a popular algorithm to adjust the interconnection weights during training. The BP algorithm, based upon the generalized delta rule, proposed by Rumelhart *et al.* (1986) is used in this

Table 1 Input parameters for computational cases

	ANN model inputs	NSSS	Mean Absolute Error (cm)
Case 1	$ra(t)$ and $WL(t)$	0.982	4.688
Case 2	$ra(t)$, $WL(t-1)$ and $WL(t)$	0.982	4.623
Case 3	$ra(t)$,$WL(t-2)$, $WL(t-1)$ and $WL(t)$	0.984	4.448
Case 4	ARMAX(1,1,1,1)	0.946	8.047
Case 5	ARMAX(2,1,1,1)	0.937	8.133
Case 6	ARMAX(3,1,1,1)	0.938	8.140

study. After learning is complete, the weights are frozen. To validate its performance testing data set is used.

Application and interpretations

The characteristics of the study area were described extensively by various researchers (Kempe *et al.*, 1978; Kadioglu *et al.*, 1997; Şen *et al.*, 1999). The lake levels data show a non-stationary behavior that can be seen clearly from Figure 1. In order to treat non-stationarity, the data considered as four different portions and then the trend is removed from those portions (Figure 4). These linear trends embedded in the historical records are detected by least square regression line which is shown on the figure. The actual data is subtracted from the trends and the rest is called residuals. Finally annual cycle which is computed by using Fourier series is removed from the detrended data.

For a one-month-ahead prediction of the Lake Van water level, several one-hidden-layer and two-hidden-layers back-propagation ANNs were trained and tested. The result is a basic three-layered network as shown in Figure 3. The smallest possible number of parameters is desired for adequate representation of the ANN model. The process of selection the most suitable model is done by trial and error.

In the application of ANNs to work on combinations of the consecutive lake level and rainfall measurements, all networks (see Table 1) were trained for specific arrangements

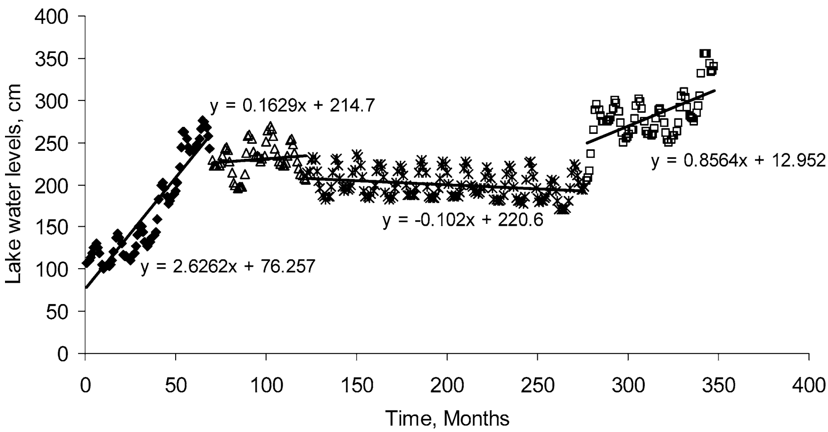

Fig. 4 Water level fluctuations and trends of Van lake

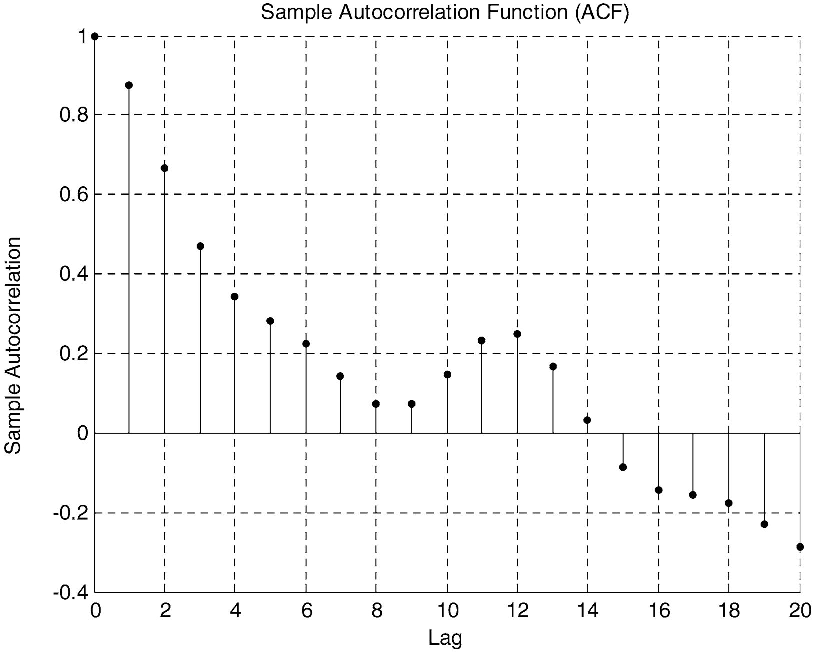

Fig. 5 Autocorrelation function plot

of rainfall amount, ra (mm), and lake levels, WL (cm). Moreover, the antecedent water levels were considered as an additional input node in Cases 2 and 3. Note that the aim is to predict the one month ahead lake level $WL(t+1)$. Hence, the model has three layers, namely, observations (recorded time series) as input layer, hidden layer as response, and the output layer as prediction.

Many studies agree that the process of training is an important aspect, and the performance of an ANN is crucially dependent on successful training. In this part, the ANN requires several thousands of epochs before training is accomplished. For training and testing parts, the data is divided into two parts. Herein, last five years (1990–1994) are left for the test (prediction) whereas the earlier part is employed for training.

The output layer had only one neuron, that is, the future lake level $WL(t+1)$. The number of neurons in the hidden layer is decided after many trials. In the trial networks, the number of hidden layer neurons was varied from 1 to 10. The configuration that gave the minimum mean relative error defined in Equation (1) is selected for each of the options.

$$(\mathrm{MRE})_i = \frac{1}{N}\sum_{i=1}^{N}\frac{|WL_{pi} - WL_{mi}|}{WL_{mi}} \times 100 \tag{1}$$

where WL_{pi} and WL_{mi} are the predicted and measured lake level values in month i and N is the total number of observations. In order to investigate the superiority of the ANN, this part is considered as a system identification problem and a comparison is made with ARMAX model. The simple ARMAX (*p, q, r*) model can be presented as,

$$WL(t+1) + a_1 WL(t) + \cdots + a_p WL(t-p) = b_1 ra(t) + \cdots b_q ra(t-q) + e(t) + \cdots c_r e(t-r) \quad (2)$$

where $WL(t+1)$ is the water level in month $t+1$, $WL(t)$ the is water level in month t, $ra(t)$ is rainfall amount in month t, a_t, b_t, and c_t are the weights of the autoregressive and moving average process respectively, the white noise, e_t, is normally distributed with zero mean and unit variance. The order of the ARMAX model that gives the highest autocorrelation is determined from Auto Correlation function (ACF) plot (Figure 5). In order to compare with ANN model, the model selected for ARMAX is (1, 1,1), which can be expressed by the following mathematical expression:

$$WL(t+1) + 0.8382\, WL(t) = 0.049 ra(t) + e(t) + 0.313 e(t-1) \quad (3)$$

The Nash-Sutcliffe Sufficiency Score (NSSS) and mean absolute error (MAE) which are defined in Equations (4) and (5) are used to evaluate the accuracy of prediction for training and testing data. NSSS is defined as the ratio of the mean square error to the variance in the observed data, subtracted from unity. It ranges from $-\infty$ to 1, with higher values indicating better agreement between the observed and predicted values.

$$\mathrm{NSSS} = 1 - \frac{\frac{1}{N}\sum_{i=!}^{N}(WL_{pi} - WL_{mi})^2}{\frac{1}{N}\sum_{i=!}^{N}(WL_{mi} - \overline{WL})^2} \quad (4)$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=!}^{N}|WL_{pi} - WL_{mi}| \quad (5)$$

The testing data was used to compare the models. The ANN and ARMAX predictions are very reasonable when compared with the recorded levels. The NSSS and MAE values are shown in the Table 1's for different cases. It can be said that ANN predictions slightly outperformed compared to ARMAX model. The predictions for the test period of five years (1990–1994) are shown in Figure 6. The superiority of ANN (Case 3) and ARMAX (Case 4) models is also reflected in the mean water level predictions for 60 months. The observed value is 303.42 cm, ANN model predicted it as 304.29 cm and the ARMAX model as 305.71 cm.

Due to the space limitations only the monthwise prediction results are given in Table 2. As seen from the table the lowest observed value 250.0 cm in November 1991 and those yielded by ARMAX and ANN are 254.57 cm and 249.96 cm respectively. On the other hand, the highest observation value is 309.0 cm in July 1992 and those predicted by ANN and ARMAX models are 303.43 cm and 305.89 cm, respectively. Both for highest and lowest observation values ANN model produce better predictions. When evaluation is considered in terms of average relative error, it is seen that ANN model exhibits more accurate results (Table 1).

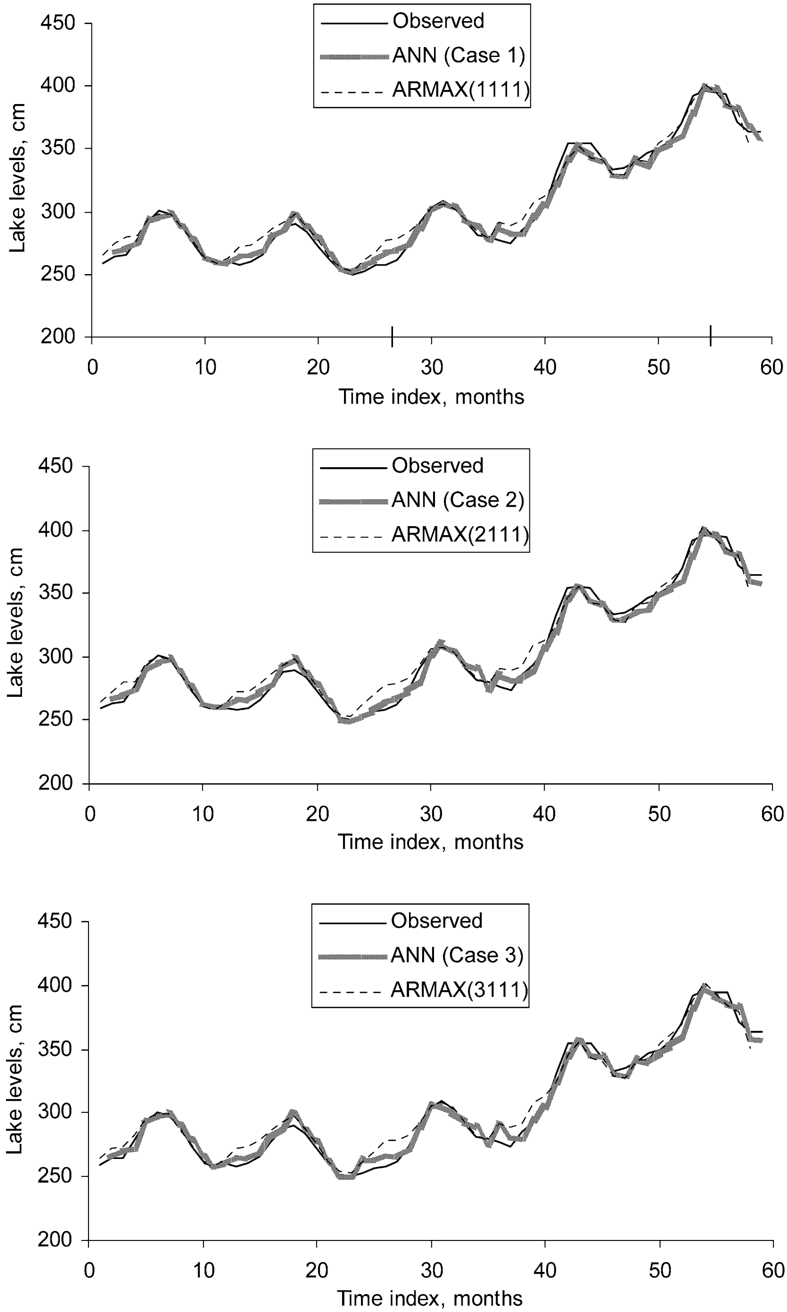

Fig. 6 Comparison of ANN and ARMAX predictions a) Case 1 and Case 4, b) Case 2 and Case 5 and c) Case 3 and Case 6 for five years 1990 and 1994

Table 2 Lake levels prediction (cm) for 1991–1992

Years	Months	Observed lake level	Predicted lake level: ARMAX(1111) (Case 4)	Predicted lake level: ANN (Case 3)	Relative errors (%): ARMAX (1111)	Relative errors (%): ANN (Case 3)
1991	January	258	263,12	266,32	1,95	3,13
	February	260	272,01	264,42	4,41	1,67
	March	266	273,26	269,52	2,66	1,31
	April	277	279,41	280,22	0,86	1,15
	May	288	286,65	288,18	0,47	0,06
	June	290	291,90	299,43	0,65	3,15
	July	284	297,79	285,84	4,63	0,64
	August	272	287,07	276,95	5,25	1,79
	September	261	276,81	262,22	5,71	0,46
	October	252	262,37	250,85	3,95	0,46
	November	250	254,57	249,96	1,79	0,02
	December	253	252,68	262,81	0,13	3,73
1992	January	257	261,02	262,63	1,54	2,14
	February	258	268,60	266,99	3,95	3,37
	March	262	277,27	265,36	5,51	1,27
	April	273	278,66	272,55	2,03	0,17
	May	291	283,29	288,62	2,65	0,82
	June	304	292,01	305,38	3,94	0,45
	July	309	303,43	305,89	1,80	1,01
	August	303	305,39	301,14	0,78	0,61
	September	291	302,31	292,83	3,74	0,63
	October	281	292,65	288,57	3,98	2,62
	November	280	284,01	274,86	1,41	1,84
	December	277	278,44	289,94	0,52	4,46
Average		**274,87**	**280,20**	**277,98**	**2,68**	**1,54**

Conclusions

In this study three different cases were considered with the network trained for different arrangements of input nodes such as past lake levels and rainfall amounts. The subsequent model provides the prediction of one month ahead lake level. It is concluded that ANN and ARMAX models are very useful for the short term predictions of the time series data. In any application, ANN system is constructed based on the training data set, which includes the first portion of the available record. The ANN models are used for the prediction purpose, provided that the values of input nodes are given. In the ANNs there are no restrictive assumptions such as linearity, normality, stationarity, ergodicity, independence of residuals, etc. The application of the methodology is presented for water level fluctuations in Lake Van, which lies in the eastern Turkey. The predictions are obtained for the five years test period. It is concluded that ANN models outperform ARMAX models. It is recommended that for more acceptable ANN models among hydrologists, some more applications should be forthcoming.

References

Altunkaynak A, Özger M, Şen Z (2003) Triple diagram model of level fluctuations in Lake Van, Turkey. Hydrol and Earth Sys Sci 7(2):235–244

Agarwal A, Singh RD (2004) Runoff modelling through back propagation artificial neural network with variable rainfall-runoff data. Water Resour Mngmt 18(3):285–300

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000a) Artifical neural networks in hydrology I: preliminary concepts. J Hydrol Engrg 5(2):115–123

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000b) Artifical neural networks in hydrology I: hydrologic applications. J Hydrol Engrg 5(2):124–137

Box GEP, Jenkins GM, Reinsel GL (1994) *Time series analysis:* Forecasting *and* control, 3rd edition, Prentice-Hall, Englewood Cliffs, NJ

Cancelliere A, Giuliano G, Ancarani A, Rossi G (2002) A neural networks approach for deriving irrigation reservoir operating rules. Water Resour Mngmt 16(1):71–88

Ellis GW, Yao C, Zhao R, Penumadu D (1993) Stress-straing modeling of sands using artificial neural networks. J Geotech Engrg 121(5):429–435

Frederick KD, Gleick PH (2001) Potential impacts on U.S. water resources, Climate change: science, strategies, & solutions, Claussen, E (ed), Pew center on global climate change, Virginia

Iruine KN, Eberthardt AK (1992) Multiplicative seasonal ARIMA models for lake Erie and Lake Ontario water levels. Water Resour Bull 28(3):385–396

Jain SK, Chalisgaonkar D (2000) Setting up stage-discharge relations using ANN. J Hydrol Engrg 5(4):428–433

Jain A, Varshney AK, Joshi UC (2001) Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks. Water Resour Managm 15(5):299–321

Kang KW, Kim JH, Park CY, Ham KJ (1993) Evaluation of hydrological forecasting system based on neural network model. Proc. 25th Congress of Int. Assoc. for Hydr. Res., International Association for Hydraulic Research, Delft, The Netherlands, 257–264

Kadıoğlu M, Sen Z, Batur F (1997) The greatest soda-water lake in the world and how it is influenced by climatic change. Ann Geophysicae 15:1489–1497

Kadıoğlu M, Sen Z, Batur F (1999) Cumulative departures model for lake-water fluctuations. J Hydrol Engrg 4(3):245–250

Kempe S, Khoo F, Gürleyik V (1978) Hydrography of lake Van and its drainage area, pp. 30–45, in the Geology of Lake Van, Degen ET, Kurtman F (eds), The mineral research and exploration institute of Turkey, Rep. 169

Kumar DN, Raju KS, Sathish T (2004) River flow forecasting using recurrent neural networks. Water Resour Mngmt 18(2):143–161

Loucks FD (1989) Modeling the great lakes hydrologic-hydraulic system. Ph D Thesis, University of Wisconsin, Madison

McCulloch WS, Pitts W (1943) A logical calculus of the ideas imminent in nervous activity. Bull Math Biophys 5:115–133

Muleta MK, Nicklow JW (2005) Decision support for watershed management using evolutionary algorithms. J Water Resour Plng and Mgmt 131(35):35–44

Sivapragasam C, Muttil N (2005) Discharge rating curve extension – A new approach. Water Resour Mngmt 19(5):505–520

Slivitzky M, Mathier L (1993) Climatic changes during the 20th century on the Laurentian Great Lakes and their impacts on hydrologic regime, NATO Adv Study Inst, Deauaille, France

Suen J, Eheart JW (2003) Evaluation of neural networks for modeling nitrate concentrations in rivers. J Water Resour Plan Manag 129(6):505–510

Şen Z, Kadıoğlu M, Batur E (1999) Cluster regression model and level fluctuation features of Van Lake, Turkey. Ann Geophysicae 17:273–279

Şen Z, Kadıoğlu M, Batur E (2000) Stochastic modeling of the Van Lake monthly fluctuations in Turkey. Theor and Appl Climatol 65:99–110

Şen Z, Altunkaynak A, Özger M (2004) Sediment concentration and its prediction by perceptron kalman filtering procedure. J Hydr Engrg 130(8):816–826

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representations by error propagation, Parallel disributed processing, Vol.1, Rumelhart DE, McClelland JL (eds), MIT Press, Cambridge, Mass, 318–362

Vaziri M (1997) Predicting caspian sea surface water level by ANN and ARIMA models. J Wtrwy, Port, Coast, Oc Engrg, 123(4):158–162