J. Agric. Sci. Technol. (2005) Vol. 7: 17-30

Comparison of Regression, ARIMA and ANN Models for Reservoir Inflow Forecasting using Snowmelt Equivalent (a Case study of Karaj)

K. Mohammadi[1*], H. R. Eslami[2] and Sh. Dayyani Dardashti[3]

ABSTRACT

The present study aims at applying different methods for predicting spring inflow to the Amir Kabir reservoir in the Karaj river watershed, located to the northwest of Tehran (Iran). Three different methods, artificial neural network (ANN), ARIMA time series and regression analysis between some hydroclimatological data and inflow, were used to predict the spring inflow. The spring inflow accounts for almost 60 percent of annual inflow to the reservoir. Twenty five years of observed data were used to train or calibrate the models and five years were applied for testing. The performances of models were compared and the ANN model was found to model the flows better. Thus, ANN can be an effective tool for reservoir inflow forecasting in the Amir Kabir reservoir using snowmelt equivalent data.

Keywords: ARIMA, Artificial neural network, Regression analysis, River flow forecasting.

INTRODUCTION

New theories concerning the human brain introduced a new approach to move our conventional digital computers on to a new computation and computer architecture. One such computational system, an artificial neural network (ANN), learns to solve a problem by developing a memory capable of associating a large number of input patterns with a resulting set of outputs or effects. The ANN develops a solution system by training on examples given to it. In this paper, ANNs were studied in the context of reservoir inflow prediction using the snowmelt equivalent in a watershed.

Inflow is important data for an optimal reservoir operation. There are several inflow forecasting methods including the time series analysis approach, rainfall-runoff modeling, and regression analysis (Hsu *et al*., 1995). Recently, ANN models have attracted increased attention due to their effectiveness and viability. While traditional models are of importance in the understanding of hydrologic processes, there are many practical situations where the main concern is with making accurate predictions at specific watershed locations (Hsu *et al*., 1995). In such a situation, using a simpler system that relates some available data to inflow may be preferred.

The Karaj river watershed, located to the northwest of Tehran, Iran, was selected to demonstrate the applicability of different methods of predicting spring inflow. The watershed has a drainage area of 850 km^2 and the average elevation is 2806 m above sea level. The monthly stream flow at Amir-

[1] Faculty of Agriculture, Tarbiat Modarres University, P. O. Box 14115-336, Tehran, Islamic Republic of Iran.
[2] Water Resources Planning Department, Jamab Consulting Engineers Company, Tehran, Islamic Republic of Iran.
[3] Jamab Consulting Engineers Company, Tehran, Islamic Republic of Iran.
[*] Corresponding author, e-mail: khuroshm@modares.ac.ir

Kabir Dam has been recorded for the period between 1970 and 1999. The precipitation and snowmelt equivalent data are collected as daily and monthly averages in the Karaj basin in five stations, respectively (Figure 1).

The main purpose of the Amir-Kabir reservoir is to provide drinking water to Tehran. A small percentage of its storage is used for irrigation. As 60 percent of the reservoir inflow occurs between April to June, prediction of the inflow in this season is very important for the reservoir operation (Figure 2). Most of the inflow in Spring is caused by the melting of the snow that falls during winter in the watershed.

The first objective of this paper was to develop an ANN model and to predict the inflow to the Karaj reservoir. The second objective was to compare the ANN model with two other methods: Auto Regressive Integrated Moving Average (ARIMA) and regression analysis. The time series models with a hydrology point-of-view have been discussed by Salas *et al*. (1980). Weeks and Boughton (1987) reviewed applications of time series models to hydrology and showed that a number of well-known hydrologic models are special cases of the ARIMA model. Multiple linear regression methods have been used widely in river flow forecasting by many researchers (Davidson, *et al*., 2002; Gorman and Toman, 1966; Lall and Bosworth, 1993; Galeati, 1990) mostly because of its simplicity and ease of use. Tokar and Johnson (1999) applied ANN models to forecast runoff as a function of daily precipitation. The application of ANN in reservoir inflow prediction and operation has been studied by Jain *et al*., (1998) who concluded that ANN is a powerful tool for input-output mapping and can be effectively used for reservoir inflow forecasting and operation.

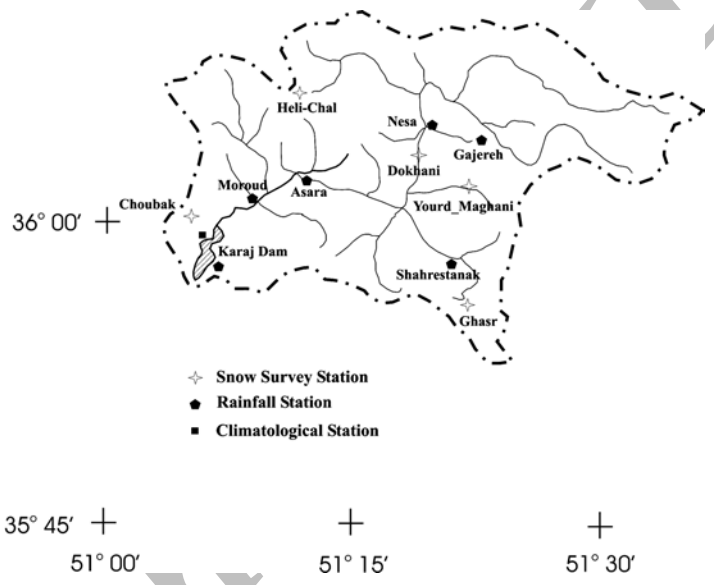

Figure 1. Station locations in Karaj watershed

ANN: An Overview

The ANN approach is based on the highly interconnected structure of brain cells. This approach is faster in comparison to its conventional counterparts, robust in noisy envi-

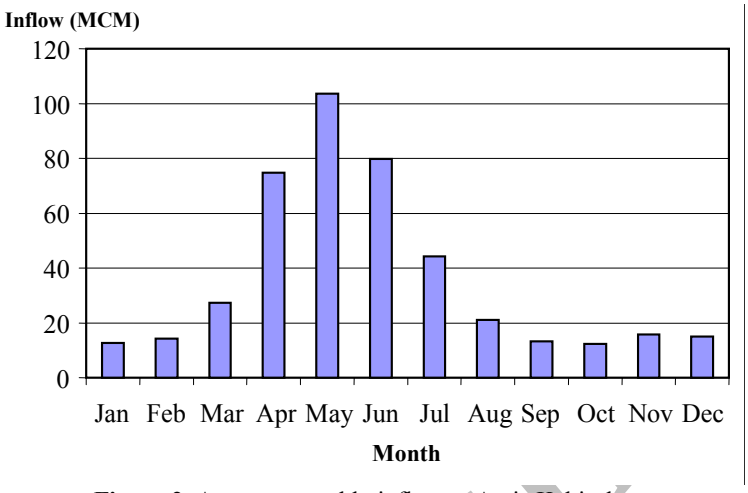

Figure 2. Average monthly inflow to Amir-Kabir dam.

ronments, flexible in terms of solving different problems, and highly adaptive to newer environments (Jain *et al*., 1999). Due to these established advantages, ANN currently has extensive applications in system engineering-related fields such as time series prediction, rule-based control, and rainfall-runoff modeling.

Early work in ANN technology was done by Rosenblatt (1962) on the perceptron. Rumelhart *et al*., (1986) and McClelland *et al*., (1986) are often credited with leading the modern renaissance in ANN technology. The addition of more complexity in the networks, specifically adding middle (hidden) layers to multi-layer perceptron networks, together with a clear explanation of the back propagation learning algorithm, overcame many of the limitations of the one or two-layered perceptron neural networks.

Since 1986, the variety of ANNs has rapidly expanded. Maren *et al*., (1990) described about 24 ANNs and Maren (1991) listed 48. Pham (1994) estimated that over

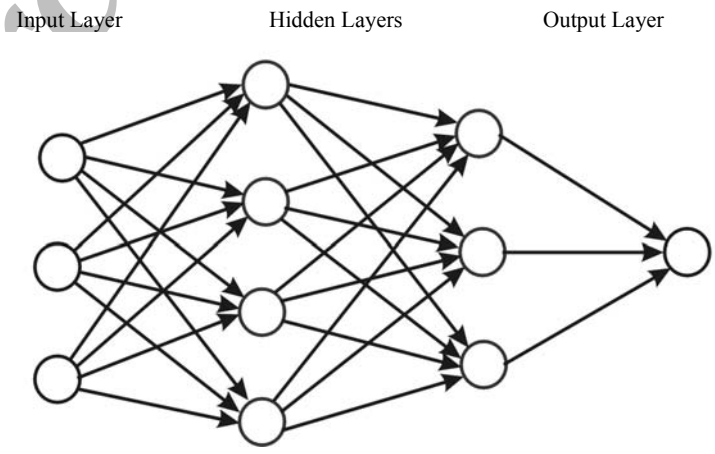

Figure 3. The basic structure of an artificial neural network.

50 different ANN types exist. There is currently a vast array of ANN applications in the cognitive sciences, the neurosciences, engineering, computer science, and the physical sciences.

The basic structure of a network usually consists of three layers: the input layer, where the data are introduced to the network; the hidden layer or layers, where data are processed; and the output layer, where the results for given inputs are produced (Figure 3).

The input values, x_i, are multiplied by weights, w_{ij}, and summed in the neuron forming $\xi_j = \sum_{i=1}^{n} X_i W_{ij}$. This result is then acted upon by an activation function, yielding the output of the jth neuron $y_j = \sigma(\xi)$, as shown in Figure 4. Only when ξ exceeds (i.e., is stronger than) the neuron's threshold limit (also called bias, *b*), will the neuron fire and become activated.

The architecture of ANN is designed by the number of layers, number of neurons in each layer, weights between neurons, a transfer function that controls the generation of output in a neuron, and learning laws that define the relative importance of weights for input to a neuron (Caudill, 1987).

Back propagation is the most commonly used supervised training algorithm (Tokar and Johnson, 1999). Werbos (1974) presented the back propagation learning algorithm for the first time but his dissertation received little attention. The algorithm was independently developed again and documented by two researchers in 1985 (Parker, 1985; Le Cun, 1985). With the development of a back propagation algorithm, the network weights are modified by minimizing the error between a target and computed outputs. In back propagation networks, the information about the error is provided backwards from the output layer to the input layer. The objective of a back propagation network is to find the weight that approximate target values of output with a selected level of accuracy.

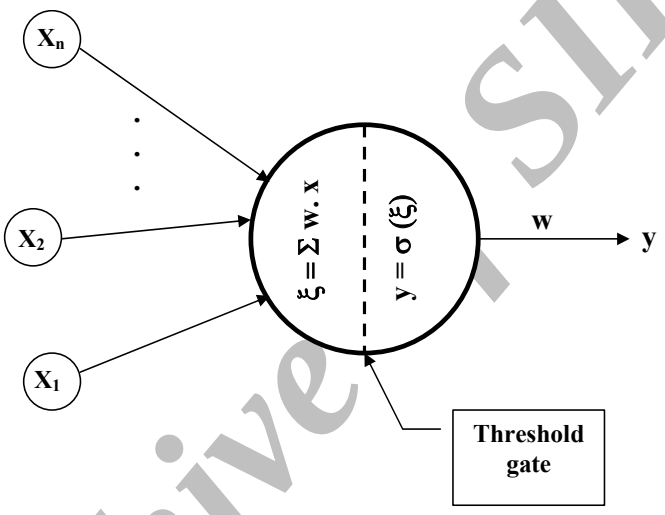

Figure 4. Signal interaction from n neurons to signal summing in the single layer perceptron.

The development of a successful ANN project constitutes a cycle of six phases, as illustrated in Figure 5 (Basheer and Hajmeer, 2003). Problem definition and formulation (phase 1) relies heavily on an adequate understanding of the problem, particularly the 'cause–effect' relationships. The benefits of

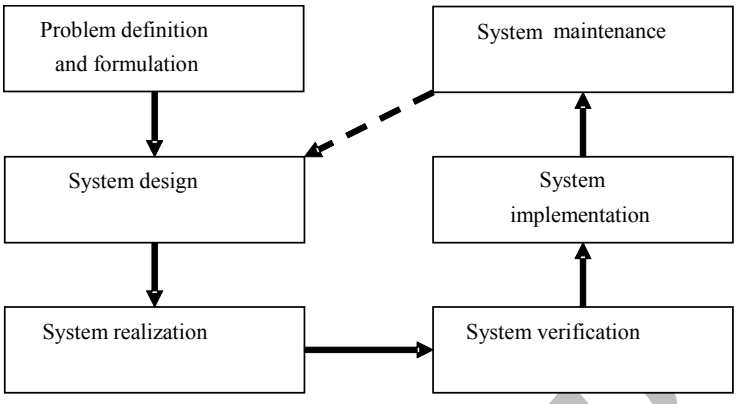

Figure 5. The phases in developing an ANN system (after Basheer and Hajmeer, 2003).

ANNs over other techniques (if available) should be evaluated before final selection of the modeling technique. System design (phase 2) is the first step in the actual ANN design in which the modeler determines the type of ANN and learning rule that fit the problem. This phase also involves data collection, data preprocessing to fit the type of ANN used, statistical analysis of data, and partitioning the data into three distinct subsets (training, test, and validation subsets). System realization (phase 3) involves training of the network utilizing the training and test subsets, and simultaneously assessing the network performance by analyzing the prediction error. Optimal selection of the various parameters (e.g., network size, learning rate, number of training cycles, acceptable error, etc.) can affect the design and performance of the final network. Splitting the problem into smaller sub-problems, if possible, and designing an ensemble of networks could enhance the overall system accuracy. This takes the modeler back to phase 2.

In the system verification (phase 4), although network development includes ANN testing against the test data while training is in progress, it is good practice (if data permits) for the 'best' network be examined for its generalization capability using the validation subset. Verification is intended to confirm the capability of the ANN-based model to respond accurately to examples never used in network development. This phase also includes comparing the performance of the ANN-based model to those of other approaches (if available) such as statistical regression and expert systems. The system implementation (phase 5) includes embedding the obtained network in an appropriate working system such as hardware controller or computer program. Final testing of the integrated system should also be carried out before its release to the end user. System maintenance (phase 6) involves updating the developed system as changes in the environment or the system variables occur (e.g., new data), which involves a new development cycle.

MATERIALS AND METHODS

Three approaches were adopted for reservoir inflow forecasting: ARIMA time series modeling, regression analysis between some hydrometeorological data and inflow, and the ANN model. Forecasts for the seasonal average of three Spring months, April, May and June, were obtained from all models and compared with the actual inflow to investigate which approach gives better predictions.

Forecasting Using Regression Analysis

Three models were selected to predict the monthly streamflow in spring (Jamab, 1997). The first one was selected by representing streamflow at the present time step i, as a function of snowmelt equivalent for watershed in winter and rainfall at time step i-1:

$$I_i = A P_{i-1} + B S_{i-1} \quad (1)$$

where I_i is the inflow to the reservoir in month i [in million cubic meters (MCM)/ month], P_{i-1} is the cumulative rainfall over the watershed using the weighted average over four stations from October to the previous month [in mm], S_{i-1} is the snowmelt equivalent again using the weighted average over five stations [in mm], A and B are constant coefficients. Snowmelt equivalent data is only available for March and April so, in order to estimate the inflow in May and June, the snowmelt equivalent in April is used.

For the second model, an additional parameter was added to the previous model, which is the temperature in the previous month:

$$I_i = A P_{i-1} + B S_{i-1} + C T_{i-1} \quad (2)$$

where T_{i-1} is the temperature in the previous month [^oC] and C is the coefficient of temperature.

To develop the third model, the inflow at time step i-1 was added to Equation 2:

$$I_i = A P_{i-1} + B S_{i-1} + C T_{i-1} + D I_{i-1} \quad (3)$$

Using 25 years of data (1970-1994), three equations were adapted for each month in Spring. Tables 1 to 3 show the resulting coefficients and regression coefficients for the calibration period. Then, these models were used to predict the inflow for the remaining 5 years of data. In these tables, R is the regression coefficient between observed and calculated inflows.

Table 1. Regression parameters between predicted inflow and precipitation and snowmelt equivalent.

Month	A	B	r
April	0.2412	0.1270	0.96
May	0.3824	0.1191	0.96
June	0.6725	0.1057	0.92

Table 2. Regression parameters between predicted inflow and precipitation, snowmelt equivalent, and temperature.

Month	A	B	C	r
April	0.2227	0.1012	1.2636	0.96
May	0.1151	0.0791	3.7211	0.95
June	0.3083	0.0541	2.3356	0.95

Forecasting with Arima Models

A time series is a set of observations generated sequentially in time. If a stationary stochastic process, a process whose parameters do not change over time, can describe the stream flow population, and if a long historic stream flow record exists, then a statistical stream flow model may be fitted to the historic flows. This statistical model can then generate synthetic sequences that reproduce selected characteristics of the historic flows. An auto regressive integrated moving average (ARIMA) method was used to model the historic flows and predict future stream flows on the basis of the past stream flows only.

The method of least squares was used to estimate the parameters. The accuracy of a forecast is best assessed by comparing the forecasts made and the values observed during the forecast periods.

The general class of ARIMA model can be written as follows (Box and Jenkins, 1976):

$$\phi_p(B)\Phi_p(B^{12})Z_t = \theta_q(B)\Theta_Q(B^{12})a_t \quad (4)$$

where ϕ_p , Φ_P , θ_q , Θ_Q are polynomials of order p, P, q, and Q, respectively, and a_t is an independent random variable series with a mean of zero and variance σ_a^2. A number of models were applied to the series and finally a mixed ARIMA (1,0,1)(0,1,1) model

Table 3. Regression parameters between predicted inflow and precipitation, snowmelt equivalent, temperature, and previous inflow.

Month	A	B	C	D	r
April	0.1117	0.0983	-0.7126	1.1232	0.98
May	0.2149	0.0371	-0.9954	1.0884	0.98
June	0.3269	0.0031	-0.3140	0.7272	0.99

was selected.

Forecasting through the ANN Model

In this study, the training of the ANN model was accomplished by a back propagation algorithm. One of the important factors in ANN models is to choose the optimal network's topology. An understanding of the topology as a whole is needed before the number of hidden layers and the number of processing elements (PEs) in each layer can be estimated.

A multilayer perceptron with two hidden layers has the power of solving any problem if the number of PEs in each layer and the training time is not constrained. This is a very important result but is only an existence proof, so it does not say how such networks can be designed. The problem left for the experimenter is to find out what is the right combination of PEs and layers to solve the problem with acceptable training times and performance (Lefebvre and Principe, 1998). Training is the process by which the free parameters of the network (i.e. the weights) find their optimal values. The weights are updated using either supervised or unsupervised learning. In this research, a supervised approach was used to train the ANN models.

Several ANN structures have been tested to obtain the best results. Table 4 shows the comparison between different model topologies. The parameters chosen as input data were snowmelt equivalent depth at five stations in the watershed, the cumulative rainfall from October to March, temperature in March, and river inflow in March. Different combinations of these input data were tested to find the best predictor model. As it is shown in Table 4, the MLP model with all eight input data consisted of an 8-4-3 layer which means 8 neurons in the input layer, seven neurons in the hidden layer and 3 neurons in the output layer has the least errors.

After the training using the 25 years of data, the model was used to predict the last 5 years, which were not included in the training process. During the test period, weights were kept constant and then river flow was estimated. Trained back propagation networks tend to give reasonable answers when presented with inputs that they have never seen.

Table 4. Topologies and structures of tested ANN models.

Model	Structure	Input data	Train			Test		
			MSE	Error	r	MSE	Error	r
MLP	5-4-3	All snow data	0.35	12.25	0.78	0.47	17.24	0.84
GFF	5-4-3	All snow data	0.39	13.49	0.76	0.47	17.24	0.85
MNN	5-4(4)-4 (4)-3	All snow data	0.56	17.04	0.66	0.50	21.24	0.84
MLP	6-4-3	All snow data and rainfall	0.27	10.96	0.84	0.72	27.07	0.66
GFF	6-4-3	All snow data and rainfall	0.32	12.47	0.82	0.55	16.94	0.84
MNN	6-4(4)-4 (4)-3	All snow data and rainfall	0.47	15.15	0.73	0.67	26.65	0.68
MLP	8-4-3	All data	0.14	7.98	0.93	0.41	17.63	0.89
GFF	8-4-3	All data	0.15	8.14	0.92	0.68	22.47	0.79
MNN	8-5(5)-4(4)-3	All data	0.21	9.56	0.88	0.67	25.74	0.81

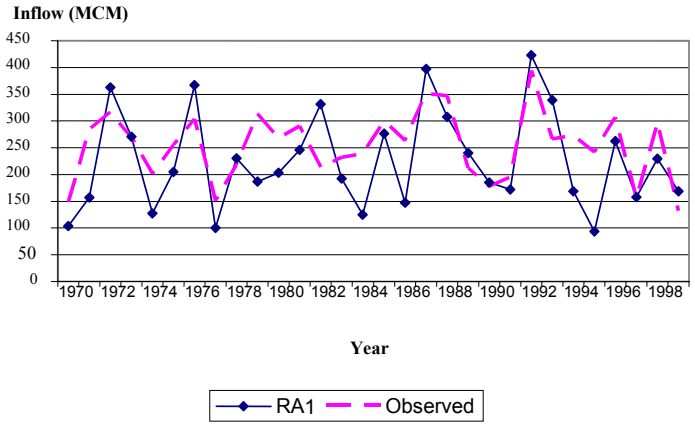

Figure 6. Comparison between the first approach of regression analysis (RA1) and observed inflow in spring season.

RESULTS

Three criteria, namely, average percentage error, average seasonal deviation, and root mean square (RMS) error between observed and calculated inflows were used to monitor the performance of the forecast models. The average percentage error is equal to:

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{ABS(Q_{obs}(i)-Q_{cal}(i))}{Q_{obs}(i)}\times 100\right] \quad (5)$$

Average seasonal deviation was calculated by:

$$\frac{1}{n}\sum_{i=1}^{n}\left[Q_{obs}(i)-Q_{cal}(i)\right] \quad (6)$$

and RMS error is:

$$\left[\frac{1}{n}\sum_{i=1}^{n}\left[Q_{obs}(i)-Q_{cal}(i)\right]^2\right]^{1/2} \quad (7)$$

where, Q_{obs} and Q_{cal} represent the observed and calculated flows, respectively, and n is the number of data. The RMS error is more pronounced with higher deviations, whereas the average percentage error is influenced by low flows. The average seasonal deviation is an unbiased interpreter of the forecast performance. Data from 1970 to 1994 was used for model calibration and training. Then, models were used to predict the spring inflow from 1995 to 1999.

The comparison between observed and computed inflows in correspondence of both calibration and validation data for the three methods of regression analysis are shown in Figures 6 to 8, respectively and Figures 9 and 10 show the comparison between observed and computed seasonal inflow values in Amir-Kabir Dam station from April to June using ARIMA and ANN methods, respectively. Table 5 shows the computed error percentages for all methods in the calibration period. It may be seen that, among the regression analysis methods, the second equation performs better with an average percentage error of 17.34 percent compared

Table 5. Calculated errors for different prediction methods in calibration period.

Error	RA1	RA2	RA3	ARIMA	ANN
AP	23.38	17.34	22.30	30.78	9.72
AD	24.76	32.46	-10.29	-10.99	-25.49
RMS	69.85	56.67	67.01	95.03	21.62

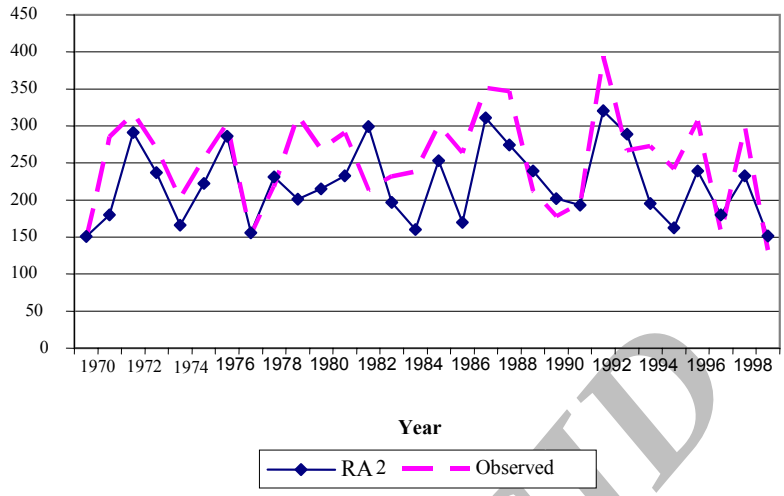

Figure 7. Comparison between the second approach of regression analysis (RA2) and observed inflow in spring season.

with 23.38 and 22.3 for the first and the third equations, respectively. The ARIMA method produced a 30.78 percent error but ANN had a significantly lower error compared with other methods. Its average percentage error was 9.72 percent. Another indicator which was calculated during calibration of the different models was a correlation coefficient between the observed and calculated data. This coefficient for RA1, RA2, RA3, ARIMA, and ANN models was 0.710, 0.665, 0.485, 0.175, and 0.937, respectively.

DISCUSSION

The objective of this study was to assess the potential application of ANN in attaining the reservoir inflow forecasting. Three different methods were used to predict the Spring inflow into the Amir-Kabir reservoir. To compare the performance of the ARIMA model, the regression analysis, and the ANN, the bar graph for average percentage error, average seasonal deviation, and RMS errors were generated using three approaches for the last five years which were not used in model fitting and training (Figures 11 to 13). The correlation coefficients for the models in the verification period were 0.545, 0.844, 0.711, 0.475, and 0.891 for RA1, RA2, RA3, ARIMA, and ANN, respectively. For thirty years of data, the errors with the ANN model are less than those for other methods. Thus, ANN can be an effective tool for reservoir inflow forecasting in the Amir Kabir reservoir.

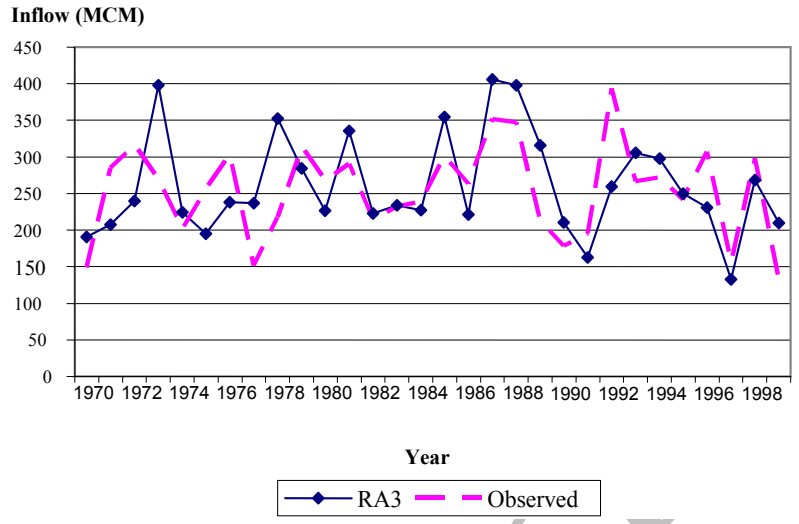

Figure 8. Comparison between the third approach of regression analysis (RA3) and observed inflow in spring season.

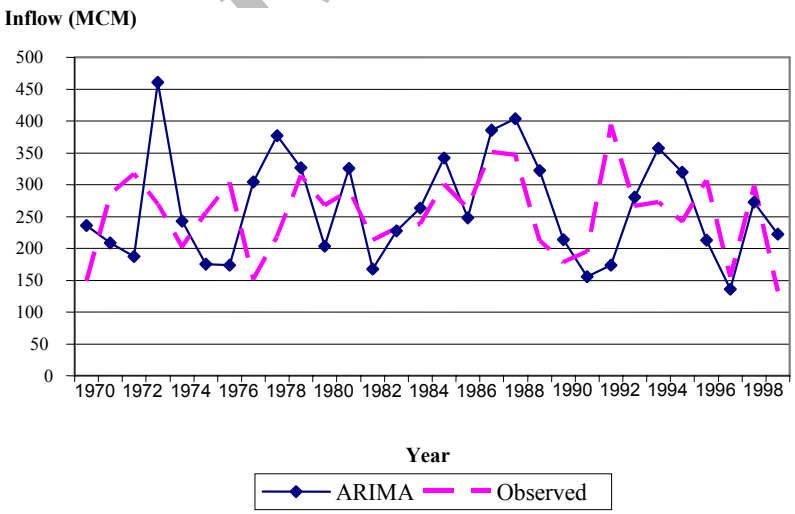

Figure 9. Comparison between the ARIMA method and observed inflow in spring season.

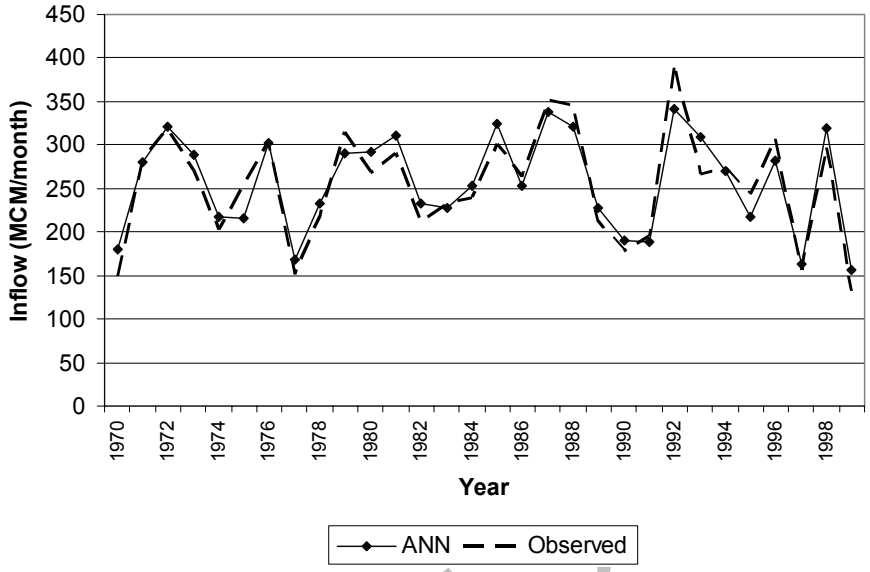

Figure 10. Comparison between the ANN model and observed inflow in spring season.

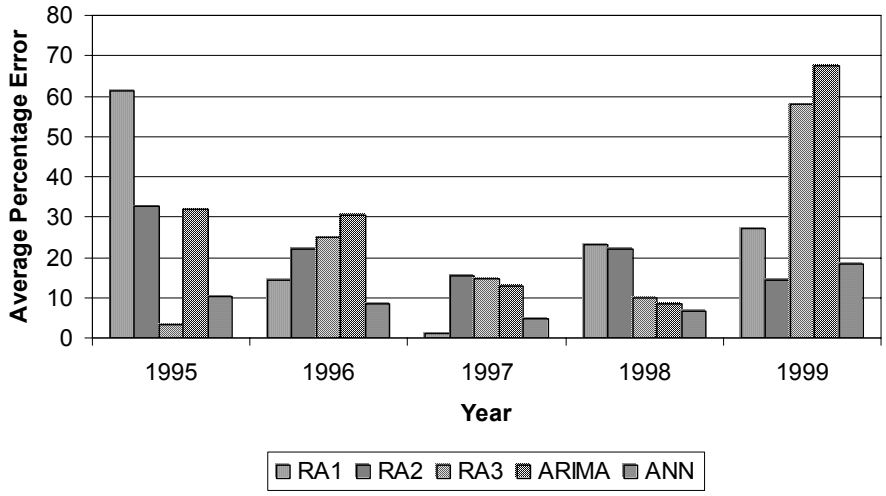

Figure 11. Average percentage error of different methods for the last five years.

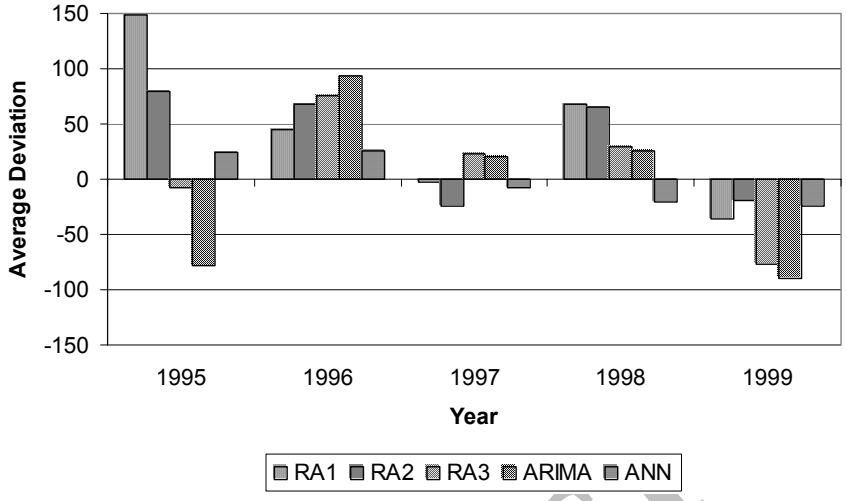

Figure 12. Average seasonal deviation of different methods for last five years.

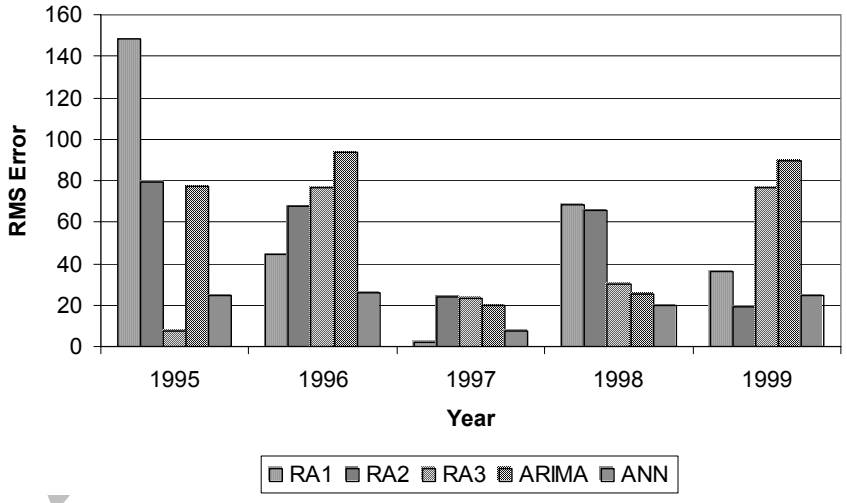

Figure 13. RMS error of different methods for last five years.

REFERENCES

1. Basheer, I. A. and Hajmeer, M. 2000. Artificial Neural Networks: Fundamentals, Computing, Design, and Application. *J. Micro. Methods*, **43**: 3-31.
2. Box, G. E. P. and Jenkins, G. M. 1976. *Time Series Analysis: Forecasting and Control*, Holden-Day Pub., Oakland, USA.
3. Davidson, J. W., Savic, D. A. and Walters, G. A. 2002. Symbolic and Numerical Regression: Experiments and Applications, *J. Inform. Sci*.
4. Caudill, M., 1987. Neural Networks Primer: Part I, *AI Expert*, December: 46-52.
5. Galeati, G., 1990. A Comparison of Parametric and Non-parametric Methods for Runoff Forecasting. *Hydrolog. Sci. J*., **35**:79–94.
6. Gorman, J. W., and Toman, R. J. 1966. Selection of Variables for Fitting Equations to Data, *Technometrics*, **8(1)**: 27-51.
7. Hsu, K., Gupta, H. V. and Sorooshian, S. 1995. Artificial Neural Network Modeling of the Rainfall-Runoff Process, *Water Resour. Res*., **31(10)**: 2517-2530.
8. Jain, S. K., Das, A. and Srivastava, D. K. 1999. Application of ANN for Reservoir Inflow Prediction and Operation, *J. Water Resour. Plan. Man.,* ASCE, **125(5)**: 263-271.
9. Jamab Consulting Engineers Company. 1997. *Climatology and Hydrology of Karaj River Watershed*. Jamab Pub., Tehran, Iran, (in Persian).
10. Lall, U., and Bosworth, K. 1993. Multivariate Kernel Estimation of Functions of Space and Time Hydrologic Data. In: *Stochastic and Statistical Methods in Hydrology and Environmental Engineering*, (Ed.) Hipel, K. Kluwer, Waterloo.
11. Le Cun, Y. 1985. Une procedure d'apprentissage pour réseau à seuil assymetrique (A learning procedure for asymmetric threshold network), *Proc. Cognit*., **85**: 599-604.
12. Maren, A. J., Harston, C. T. and Pap, R. M. 1990. *Handbook of Neural Computing Applications*, Academic Press, San Diego, Calif.
13. Maren, A. J. 1991. A Logical Topology of Neural Networks. In*: Proceedings of the Second Workshop on Neural Networks*, WNN-AIND 91.
14. McClelland, J. L., Rumelhart, D. E. and the PDP Research Group, 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 2., MIT Press, Cambridge, Mass.
15. Parker, D. B. 1985. *Learning-logic, Tech. Rep*. TR-47, Center for Computer Research in Economics and Management Science, MIT.
16. Pham, D. T. 1994. Neural Networks in Engineering. In: *Applications of Artificial Intelligence in Engineering IX, AIENG/ 94, Proceedings of the 9th International Conference*. (Eds.), Rzevski, G. *et al*. Computational Mechanics Publications, Southampton, UK. pp. 3–36.
17. Rosenblatt, F. 1962. *Principles of Neurodynamics: Perceptrons and the Theory of Brain mechanisms*, Spartan, Washington, D. C.
18. Rumelhart, D. E., McClelland, J. L. and the PDP Research Group. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, MIT Press, Cambridge, Mass.,
19. Salas, J. D., Deulleur, J. W., Yevjevich, V. and Lane, W. L. 1980. *Applied Modelling of Hydrologic Time Series.* Water Resources Publications, Littleton, Colo.
20. Tokar, A. S. and Johnson, P. A. 1999. Rainfall-Runoff Modeling Using Artificial Neural Networks, *J. Hydrol. Eng.*, **4(3)**: 232-239,
21. Weeks, W. D. and Boughton W. C. 1987. Tests of ARMA Model Forms for Rainfall-Runoff Modeling, *J. Hydrol*., **91**: 29-47.
22. Werbos, P., 1974. Beyond Regression: New Tools for Prediction and Analysis in Behavioral Sciences. Ph.D. dissertation, Harvard Univ., Cambridge, Mass.

کاربرد روش شبکه عصبي مصنوعي در پيشبيني جريان ورودي ناشي از ذوب برف به مخزن سد اميرکبير

ک. محمدي, ح. ر. اسلامي و ش . ديانـي دردشتي

چکيده

سه روش مختلف براي پيشبيني جريان در فصل بهار به مخزن سد اميرکبير که در نزديکي تهران واقع شده است بکار رفت. جريان ورودي در فصل بهار در حدود 60 درصد جريان سالانه مي‌باشد. با استفاده از مدل شبکه عصبي مصنوعي جريان ورودي پيشبيني گرديد و با دو روش ديگر يعني مدل سري زماني ARIMA و مدل همبستگي آماري بين جريان ورودي و بعضي از پارامترهاي هيدروکليماتولوژيکي حوزه مقايسه شد. با استفاده از 25 سال آمار مشاهده شده مدلهاي مذکور واسنجي گرديده و از 5 سال ديگر آمار براي صحتيابي استفاده شد. نتايج نشان داد که روش شبکه عصبي مصنوعي کارايي بهتري نسبت به ساير روشها دارد.