Stochastic modeling of Karasu River (Turkey) using the methods of Artificial Neural Networks

İbrahim Can
Atatürk University, Civil Engineering Department, 25240 Erzurum, Turkey

Cahit Yerdelen
Atatürk University, Civil Engineering Department, 25240 Erzurum, Turkey

Ercan Kahya[1]
Istanbul Technical University, Civil Engineering Department, Hydraulic Division, 34469 Maslak Istanbul, Turkey

Abstract: Engineering project design, environmental impact analysis and water resources problems, particularly hydrological simulation and forecasting often require the estimation of streamflow. Simulation and forecasting techniques of streamflow time series have been applied widely for determining the dimensions of hydraulic works for risk assessment in urban water supply systems and irrigation, optimal operation of reservoir systems, optimization hydroelectric production, and so on. The present study uses two different artificial neural network (ANN) approaches for the stochastic modeling of the monthly mean streamflow of Karasu River located in the Upper Euphrates River Basin in the Eastern Anatolia part of Turkey. The ANN approaches used in this study are well known one-hidden layer feed forward back propagation (FFBP) architecture trained with the error back propagation and a variant of radial basis networks called Generalized Regression Neural Networks (GRNN). The study shows that the GRNN performs very well in the prediction of the hydrological time series. The performance of the GRNN is also found to be comparable with that of FFBP.

1. Introduction

Modeling plays a significant role in the field of hydrological sciences. They are useful in capturing and understanding inherent high nonlinearity in both spatial and temporal domains. Over the past decades, various data-driven techniques have been adopted for modeling of hydrological time series. Key examples of such applications can be found in Salas et al. (1985), Porporato and Ridolfi (1997) and Zhang and Govindaraju (2000). Simulation and forecasting techniques of streamflow time series allow practitioners and planners to explore possible realistic future scenarios of a given water resources system (Salas et al., 1980; Koutsoyiannis, 2000).

The objective of the present study is to compare the performance of well known one-hidden layer feed forward backpropagation (FFBP) architecture

1 Assoc. Prof., Hydraulic Division
Civil Engineering Department
Istanbul Technical University
34469 Maslak Istanbul, Turkey
Tel: + 90 (212) 285-3002
e-mail: kahyae@itu.edu.tr

trained with the error back propagation and a variant of radial basis networks, so-called Generalized Regression Neural Networks (GRNN). As a case study, we considered the monthly streamflow values of the Karasu River located in the Upper Euphrates River Basin in the Eastern Anatolia region in Turkey.

2. Description of study area

Karasu River is located in the Upper Euphrates River Basin in the eastern Turkey. The location of the Karasu is shown in Figure 1. Karasu River with a drainage basin area of 2886 km^2 is elevated at 1,675 m above sea level. In this study, monthly mean flow data measured at the gauging station 2154 on Karasu River are used for the ANN modeling. The length of flow records is 32 years (or 384 months) with an observation period between 1969 and 2000. The flow data consists of hydrological years, starting at October and extending to September of the following year.

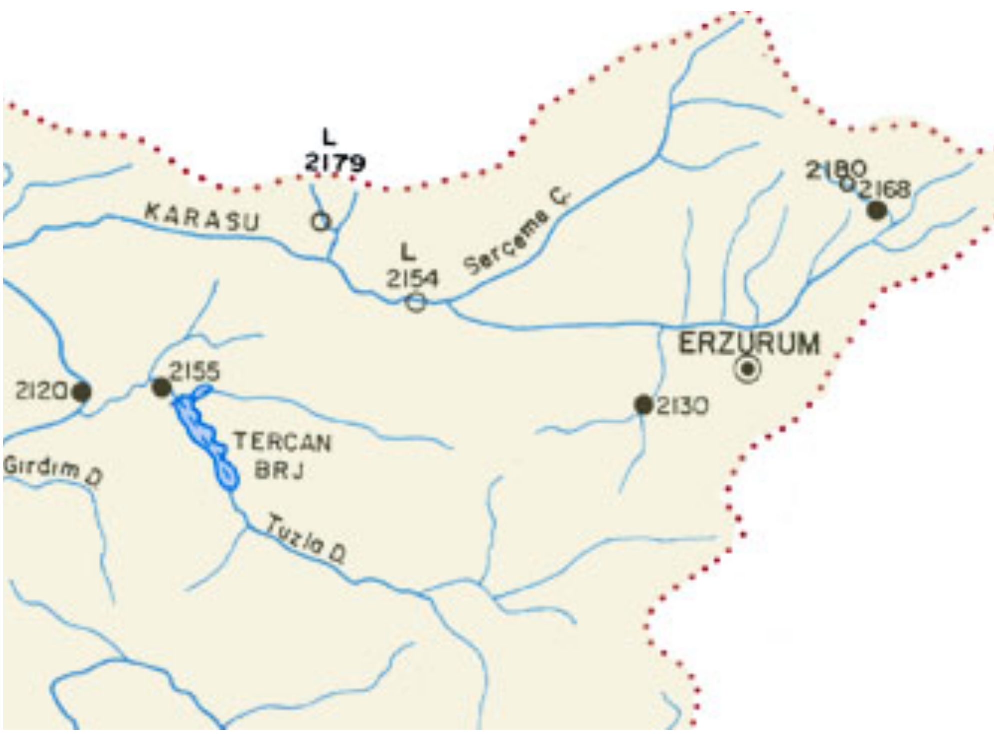

Figure 1. Location of Karasu River in the Upper Euphrates River Basin.

The Upper Euphrates River Basin was recently categorized in the hydrologically homogeneous region “Number 2” defined by Kalaycı (2003) who applied the rotated principal component analysis on a nationwide streamflow data set. Similarly Demirel (2004) used a different approach, cluster analysis, for streamflow regionalization and found 8 homogeneous regions in Turkey. The basins under consideration takes place in his Region H. Readers are referred to Ünal et al. (2003) for climate features of the basin.

3. Data preprocessing

The modeling technique applied is essentially data driven, i.e. there is no preconceived notion about the existing relationships between the variables. For efficient operation of such models, a priori analysis on data conditioning, normalizing, and scaling of the variables must be undertaken. As a first step, skewness was removed from the original records by the transformation given by

$$X_{v\tau} = \log(Q_{v\tau}) \qquad (1)$$

where $Q_{v\tau}$ is the monthly mean flow for month τ (τ=1,..., 12); year v (v =1, ..., N); and N is the number of years of record of the series.

To account for the periodicity, the resulting series after transformation through eq. (1) were standardized to improve learning efficiency and overall operation of the ANN (Salas et al., 2000). The standardization was applied on a monthly basis using the following equation.

$$Y_{v\tau} = \frac{X_{v\tau} - \overline{X}_{\tau}}{S_{\tau}} \qquad (2)$$

where $\overline{X}_{\tau}$ and S_{τ} are the sample mean and standard deviation of the normalized flows for month τ, respectively; and $Y_{v\tau}$ is the standardized value for month τ and year v.

This procedure helps to avoid internal numerical instabilities of the ANN learning process and operation (Lapedes and Farber, 1988; Govindaraju, 2000).

4. Artificial Neural Network

Artificial neural networks are mathematical models of human cognition, which can be trained to perform a specific task based on available experiential knowledge (Govindaraju, 2000). They are typically composed of three parts: inputs, one or many hidden layers, and an output layer. Hidden and output neuron layers include the combination of weights, biases, and transfer functions. The weights are connections between neurons while the transfer functions are linear or nonlinear algebraic functions. When a pattern is presented to the network, weights and biases are adjusted so that a particular output is possible to obtain.

Backpropagation is the generalization of the least mean square (LMS) algorithm to multilayer networks and nonlinear differentiable transfer functions. The multilayer feedforward network is the most used architecture of backpropagation. Feedforward networks typically consist of one or more hidden layers of nonlinear neurons followed by a layer of linear neurons. The GRNN, a variant of radial basis networks, has a radial basis layer in the first layer and a special linear layer in the second layer. Readers are referred to Demuth and Beale (1997) and Govindaraju and Ramachandra Rao (2000) for further explanations regarding the radial basis algorithms. Readers who are not familiar with the neural networks may be referred to Lippmann (1987) or some introductory books, such as Rumelhart et al. (1986) for the details.

5. Results and Conclusions

In general, the architecture of a multilayer feed forward neural network has three layers. The configuration giving the minimum mean squared error (MSE) and maximum correlation was selected for each of the combinations. In training of the FFBP networks, the hyperbolic tangent was used as an activation function. The networks trained and tested using ANN toolbox of the MATLAB software. For training the networks, flow values from October 1969 to September 1985 were selected, whereas the remainder of the dataset was used for testing the models. The number of neurons in the hidden layer was 1 to 5 and one neuron in the output layer. The 2-neuron network was selected since no improvement was made to the network when using more hidden neurons. In the input layer, a node represents an input variable. Number of past values derives from a previous exploratory phase, in which several numbers of past values (lag-intervals) were tested. When the number of lag-intervals was higher than one lag, it is decided that no significant improvement of the model performance was achieved. In this way, one lag is found to be the most appropriate number in this study. The input node in the networks represents a flow at day t-1 (Q_{t-1}) and the output represents a flow (Q_t) to be predicted. The GRNN networks are very sensitive to spread. The optimum value of spread must be obtained by trial and error. The GRNN network was constructed with a spread constant of 0.2.

In the present study, the performance of FFBP networks and GRNN in modeling monthly streamflow time series was our concern. The computed autocorrelation functions revealed that residuals are serially independent for the both models (Figures 2 and 3). Plots showing the measured and predicted streamflow values using FFBP networks and GRNN are shown in Figures 4 and 5, respectively. Finally the scatter plots of the observed versus predicted flow drawn according to the FFBP networks and GRNN are shown in Figures 6 and 7, respectively. These diagrams resemble, to a great extent, each other as their respective determination coefficients appear to be too close. According to the performance measure of RMSE, the FFBP and GRNN models behave very much alike (Figures 4 and 5).

In conclusion, the results of this study demonstrate that the performance of the FFBP networks and GRNN are comparable in terms of modeling monthly streamflow of Karasu River.

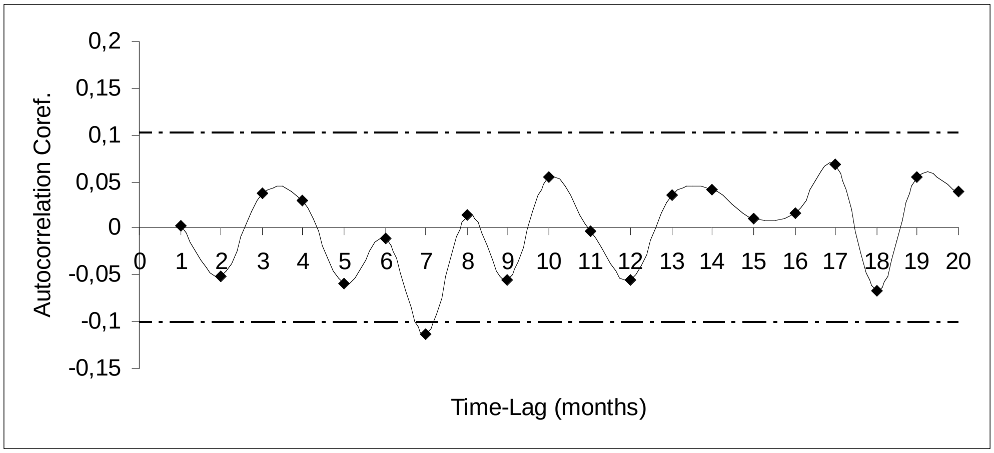

Figure 2. Autocorrelation function of residual series with the FFBP networks.

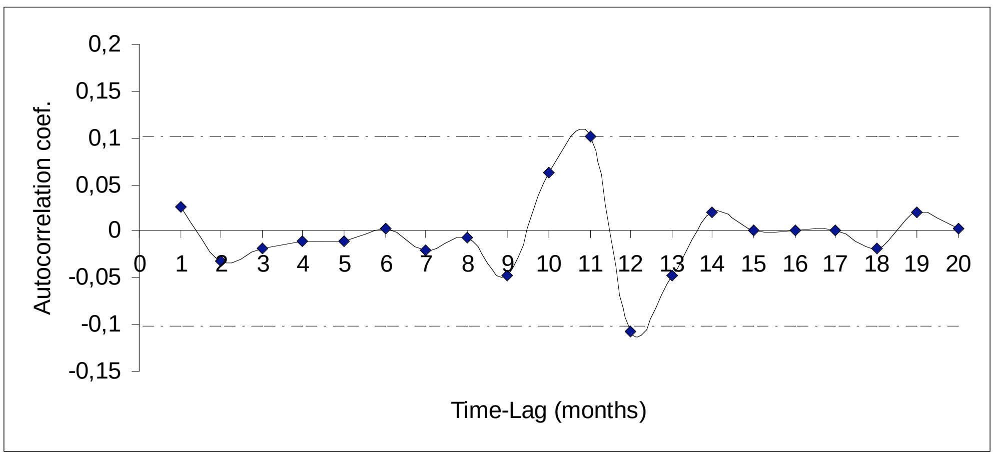

Figure 3. Autocorrelation function of residual series with the GRNN networks.

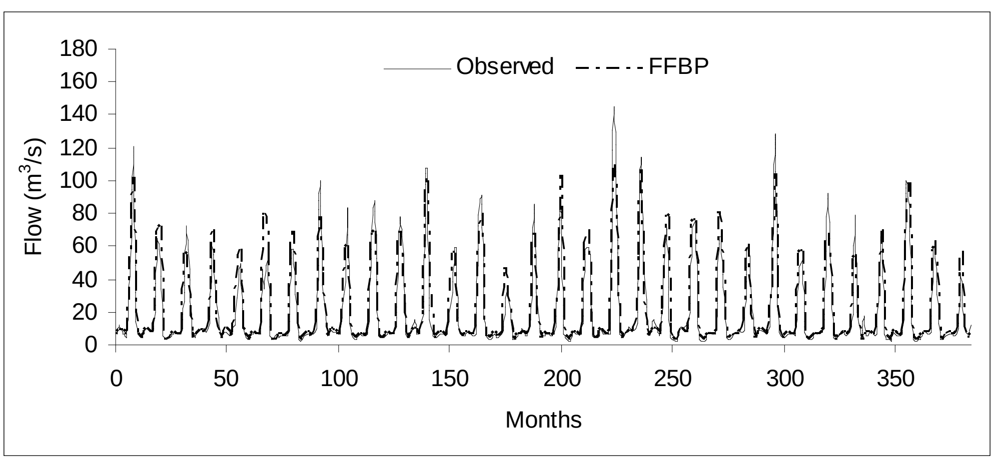

Figure 4. Observed and predicted monthly flow with the FFBP networks.

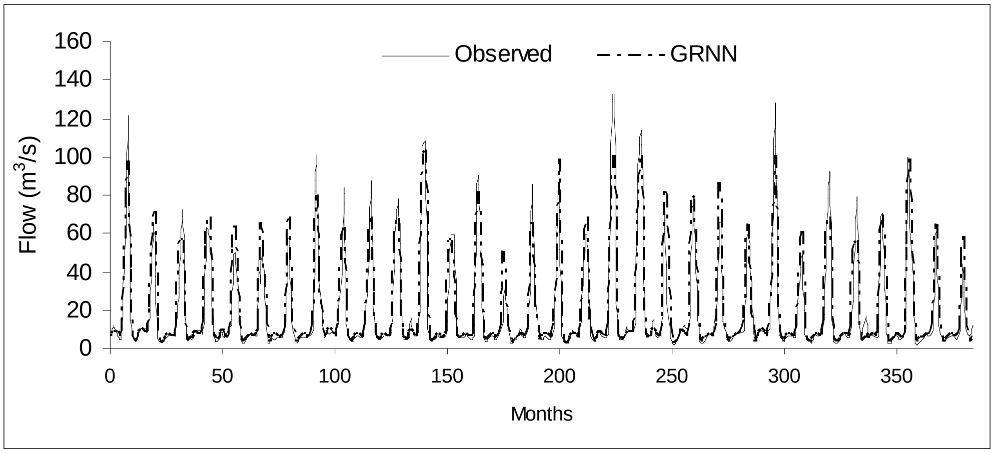

Figure 5. Observed and predicted monthly flow with the GRNN networks.

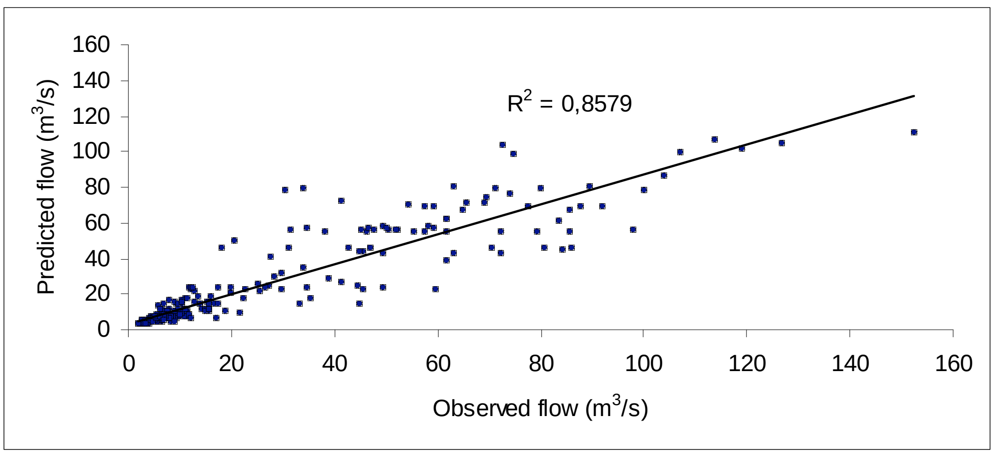

Figure 6. Scatter plot of the observed versus predicted flow with the FFBP networks.

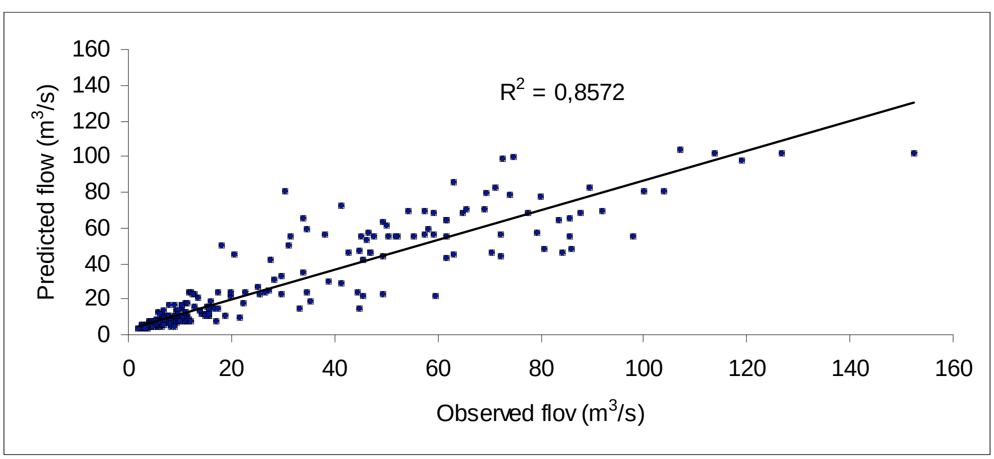

Figure 7. Scatter plot of the observed versus predicted flow with the GRNN networks.

References

Demirel, M. C., 2004: *Cluster analysis of streamflow data over Turkey.* M.Sc. Thesis, Istanbul Technical University, Istanbul.

Demuth, H., and M. Beale, 1997: Neural Network Toolbox for User's Guide. The MathWorks.

Govindaraju, R. S., 2000: Artificial neural networks in hydrology I: Preliminary concepts. *Journal of Hydrologic Engineering*, **5**, 115-123.

Govindaraju, R. S., and A. Ramachandra Rao, 2000: Artificial Neural Networks in Hydrology. Kluwer Academic Publisher.

Kalaycı, S., 2003: *Analysis of streamflow variability in Turkish rivers*. PhD Thesis (in Turkish), Selçuk University, Graduate School of Natural and Applied Sciences, Department of Civil Engineering, Konya, Turkey, 124.

Koutsoyiannis, D., 2000: A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series. *Water Resources Research*, **36**, 1519–1533.

Lapedes, A., and R. Farber, 1988: How neural net works. In: Neural information processing systems, D. Z. Anderson (Ed.), American Institute of Physics, New York, 442–456.

Lippmann, R. P., 1987: An introduction to computing with neural nets. IEEE ASSP Mag., **4**(2), 4-22.

Porporato, A., and L. Ridolfi, 1997: Nonlinear analysis of river flow time sequences, *Water Resources Research*, **33**(6), 1353-1367.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams, 1987: Learning internal representations by error propagation. Parallel distributed processing: explorations in the microstructure of cognition; Vol. 1, D. E. Rumelhart and J. L. McClelland, eds., MIT Press, Cambridge, Mass., 318-362.

Salas, J. D., J. W. Delleur, V. Yevjevich, and W. L., Lane, 1980: Applied modeling of hydrologic time series. Water Resources Publications, Littleton, Colorado.

Salas, J. D., G. Q. Tabios III, and P.Bartolini, 1985: Approaches to multivariate modeling of water resources time series. *Water Resources Bulletin*, **21**(4), 683-708.

Salas, J. D., M. Markus, and A. S. Tokar, 2000: Streamflow forecasting based on artificial neural networks. In: Artificial neural networks in hydrology, R. S. Govindaraju and A. R. Rao (Eds.), Kluwer, Dordrecht, The Netherlands, 23–51.

Zhang, B., and S. Govindaraju, 2000: Prediction of watershed runoff using Bayesian concepts and modular neural networks. *Water Resources Research*, **36**(3), 753-762.