Indian Journal of Fundamental and Applied Life Sciences ISSN: 2231– 6345 (Online)
An Open Access, Online International Journal Available at www.cibtech.org/sp.ed/jls/2016/02/jls.htm
2016 Vol. 6 (S4), pp.52-58/Fereydooni and Kasmayi
Research Article

EVALUATION OF THE ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM (ANFIS) AND STOCHASTIC MODELS (ARMA-ARIMA) IN SIMULATION OF FLOW RATE (CASE STUDY: WATERSHED OF SHAFAROUD)

***Mehrdad Fereydooni and Payvand Ebrat Kasmayi**
Department of Civil Engineering, Larestan Branch, Islamic Azad University, Larestan, Iran
**Author for Correspondence*

ABSTRACT
Estimating flow rate in the river can have an important economical role due to its effect on the management of water resources. We can calculate circular (district) flow rate by different methods. Any of these methods has its related advantages and disadvantages. In the current study, we organize an image of flow rate using adaptive neuro-fuzzy inference system and stochastic models. At first, we have calculated statistic data that include rate, participation, evaporation, temperature using hydrometric and synoptic stations in the study area. Then, we synchronized all of the statistics and used 80% of data for network education and 20% of data for model test. Evaluation criteria to select the best models are correlation coefficient (R^2) and root mean square error (RMSE). The results indicate that adaptive neuro-fuzzy networks have greater ability to simulate flow rate in the catchment river of Shafaroud.

Keywords: *Flow Rate, Neuro-Fuzzy System, Time Series, ANFIS, ARMA, ARIMA, Shafaroud*

INTRODUCTION
As it is clear, hydrological processes act non-linearly due to their diversity and are variable and unpredictable in terms of time and space that can be cited to relationship rainfall-runoff process that is the most complex hydrological processes.
Generally, the phenomenon of rainfall-runoff and forecasting river flow is done in experimental emprical models, Physical Model and Computational Model. Using adaptive neuro-fuzzy networks and time series can contribute to simulate flow rate. Application of time series in hydrology started from 4 decades ago and the models of Box and Jenkins reached their peak.
Of numerous conducted studies in this field, time series modeling of the mean flow using adaptive network based fuzzy inference system (ANFIS), artificial neural networks (Nayak *et al.,* 2004) to compare the performance, ANFIS and regression-based methods in prediction the water level of the river (Chau *et al.,* 2005), long-term runoff prediction using artificial neural networks and fuzzy inference system (Araqinezhad and Karamooz, 2005) and prediction river flow using ANFIS and artificial neural networks can be mentioned (Firat, 2007). In all these studies the models which were based on artificial intelligence had an acceptable performance in predicting the hydrologic variables under investigation. Moreover, in the research by Yurekli and Kurunc (2005) the performance of Thomas-Fiering stochastic model and auto-regressive integrated moving average (ARIMA) model in prediction monthly dehydration were compared with each other and finally ARIMA model was recognized as the more appropriate one between these two models.
There is much evidence indicating that artificial intelligence-based models due to their non-linear nature have a better performance in simulation and prediction of hydrologic variables compared to the common linear stochastic models (ASCE, 2000a,b). However, it has been seen in many cases that common stochastic models indicated a better or similar performance compared to artificial intelligence-based models in predicting the random variables (Gorr *et al.,* 1994; Maier and Dandy, 1996; Lewis and Ray, 2002).
Considering the fact that modeling of hydrologic variables depends on the type of variable under study as well as the area of the study, no one can certainly comment on the appropriate model (common stochastic or artificial intelligence based) prior to conducting a study (Mishra *et al.,* 2007).

The Study Area

The location of shafaroud river is in Iran, Guilan province and Rezvaznshahr city. The drainage basin of the river is a part of the Caspian Sea drainage basin. This basin with an area equivalent to 345 Km² is located in the west of Gilan with 48-630' E 48-41' W longitude and 37-25' S 37-34' N latitude. The maximum and minimum height of the basin is 2887 and 63 meters above the sea level, respectively.

It should be noted that the data in this study the obtained from the data of hydrometric and synoptic stations in Shafaroud include the flow rate (Q), precipitation (P), temperature (T) and evaporation (E).

Average precipitation in this area has been 1325 mm over a 19-years period (1995-2014), and precipitation occurs even in the warmest months of the year. The lowest precipitation is usually in June which is equal to 52.58 mm and the highest is in September which is equal to 207.34 mm. The average temperature over a 19-year period (1995-2014) is 15.6 °C, the minimum average is 6.67 ° C in January and the maximum average is 25.68 ° C in July. The average total annual evaporation is equal to 779.82 mm, the maximum amount of which has been observed in July.

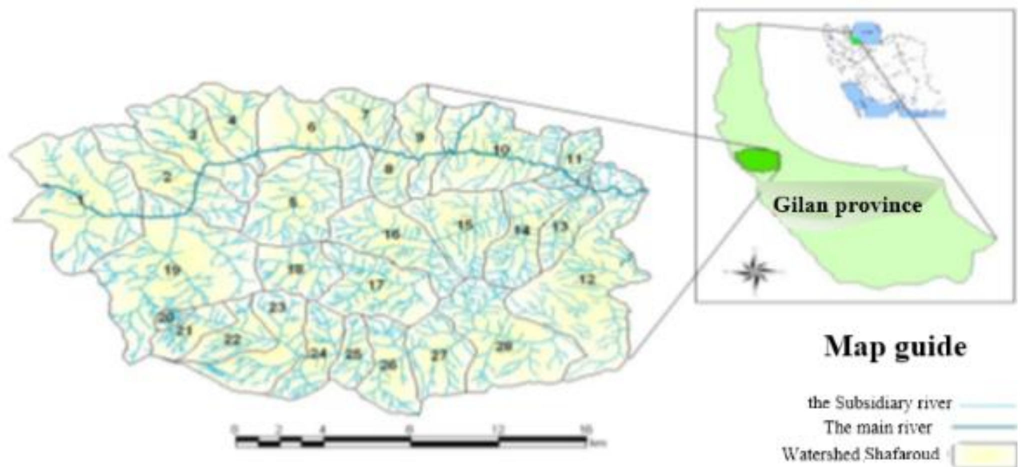

Figure1: Shafaroud River Drainage Basin in Guilan Province

Adaptive Neuro-Fuzzy Inference System (ANFIS)

Fuzzy - neural models which were expanded in 1993 by Zhang, combine fuzzy logic with artificial neural networks to facilitate the process of learning and adaptation. The most famous of these methods is adaptive neuro-fuzzy inference system (ANFIS). ANFIS has good capabilities in training, construction and classification, and also has the advantage that allows the extraction of fuzzy rules from numerical data or expert knowledge and makes a rule – basis comparatively. In addition, it can adjust turning of complex human intelligence into fuzzy systems. The main problem of ANFIS prediction model is a relatively high need of time for training the structure and determining the parameters. To simplify, we assume that the targeted inference system has two inputs of x and y and an output of z. For a first-order Takagi-Sugeno fuzzy model, a set of model law with two fuzzy IF-THEN rules can be stated as follows:

• First rule: if x is equal to A_1 and Y is equal to B_1, then $Z_1 = p_1 x + q_1 y + r_1$

• Second rule: if x is equal to A_2 and Y is equal to B_2, then $Z_2 = p_2 x + q_2 y + r_2$

Where A_1, A_2 and B_1, B_2 are membership functions for inputs x and y, respectively, and p_1, q_1, r_1, p_2, q_2 and r_2 are linear parameters in the lower section of first-order Sugeno-Takagi fuzzy model. The structure of ANFIS consists of five layers (Figure 2)

• The first layer, Input Nodes: every node of this layer produces membership values that belong to each suitable fuzzy sets using the membership function.

$$O_{1,i} = \mu_{A_i}(x) \qquad \text{for } i = 1, 2$$

$$O_{1,i} = \mu_{B_{i-2}}(x) \qquad \text{for } i = 3, 4$$

where x and y of non-fuzzy inputs to nodes i and Ai and Bi (small, large, and...), are linguistic tags that are defined by proper membership functions μBi and μAi, respectively. Here, Gaussian and bell-shaped fuzzification instruments are usually used. The parameters of the membership functions, which are known as basic parameters in this layer, should be identified.

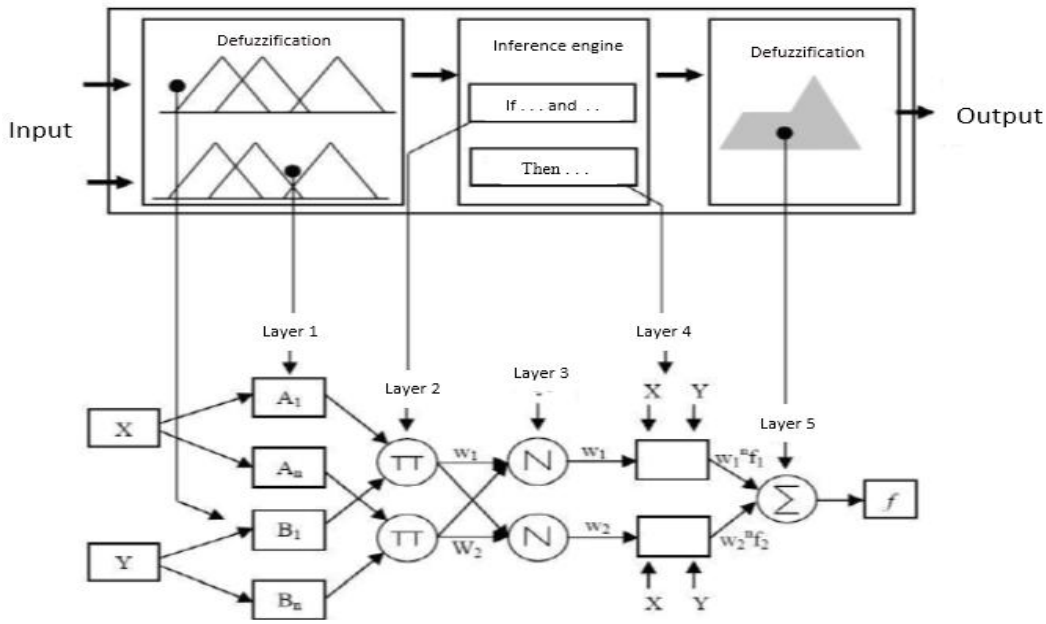

Figure 2: A Sample of ANFIS Model Structure

• The second layer, Rule Nodes: In the second layer, the operator "AND" is used in order for output (Firing Strength), that represents the first part of the law to be achieved. The degree to which the first phase of a fuzzy law has been met, is called the firing strength which shapes its output function. Hence, the output O2, and k of the layer are resulted from multiplication of grades related to the first layer. In Layer 2, input signals to circular nodes are multiplied in each other, and the output that reflects the performance strength or weight of a law, is calculated as follows:

$O_{2,k} = \mu_{A}(x) + \mu_{B}(y)$ k = 1, . . ., 4   i = 1 , 2 ; j=1 ,2

• The third layer, Average Nodes: The main purpose of the third layer is to determine the ratio of each firingstrength of ith of the law to the sum of all firing strength of the laws. As a result, it is obtained as normalized firingstrength:

$O_{3,i}=\bar{w}_i = \frac{w_i}{\sum_{K=1}^{4} w_k}$ i = 1 , . . . ,4

• The fourth layer, Consequent Nodes: the function calculates the node of the fourth distribution layer of ith law to the total output and defines it as follows:

$O_{4,i} = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i)$ i = 1 , . . . ,4

Where $\bar{w}_i$ is the output of ith node from the previous layer. {Pi, qi, ri} are the coefficients of the linear combination as well as the series of parameters of the lower part of Takagi - Sugeno fuzzy model.

• The fifth layer, Output Nodes: this single node calculates total output by adding all input signals. Thus, in this layer, defuzzification process changes results of each fuzzy rule to defuzzified output.

$O_{5,i}= \bar{w}_i f_i= \frac{\sum_{i=1}^{4} w_i f_i}{\sum_{i=1}^{4} w_i}$

In adaptive neuro-fuzzy inference system, firstly, the structure of a model with specific parameters, which is proportional to inputs, membership degree, and rules and functions of output membership degree, is selected and then a part of the available data is selected as input-output that can be used for training this system. In training stage, through modifying parameters of membership degree based on acceptable error rate, the model parameters become close to the actual values. For utilizing neuro-fuzzy system in MATLAB software package, two methods of grid partitioning and sub-clustering are considered.

Stochastic Models

Stochastic Time Series ARMA Modeling

Another method is using time series ARMA model which is an important parametric family of time series. The model uses past observations and disturbances in order to predict in time. This method requires fewer parameters in modeling than the neural network. ARMA model with parameters p and q is generally expressed as follows.

$$\mathrm{x}_t - \emptyset_1 \mathrm{x}_{t-1} \dots ..$$

Where Xt-i is observations of the flow rate level at time t-i, Zt-i is SID noise at time t-i, and $\emptyset_i, \emptyset_\mathrm{i}$ are model coefficients.

Stochastic Time Series ARIMA Modeling

Two general forms of ARIMA models include non-seasonal ARIMA (p, d, q) and multiplicative-seasonal ARIMA (P, d, q) (P,D,Q)* where q and p, are parameters of autoregressive and non-seasonal moving average, and P and Q are parameters of autoregressive and seasonal moving average, respectively. The other two parameters, meaning, d and D, are differential parameters for making the time series static. Differential operator used for dynamic time series are $\Delta = 1 - B$ (B is reverse mutation operator) and $\Delta^d = (1 - B)^d$ which is for seasonal differentiation. This form of the non-seasonal ARIMA models is written as follows:

$Z_t = \Phi(B)(1 - B)^d Z_t = \Theta(B) a_t$

Where Zt is observed series, Φ (B) is polynomial ranking of p, and θ (B) is polynomial ranking of q. For seasonal time series that are often cyclical, seasonal differencing is used and here we will have multiplicative-seasonal model:

$\varphi_p(B)\Phi_p(B^S)\Delta^d \Delta_s^D (z_t - Z) = \Theta_q (B)\, \Theta_Q (B^S)\, a_t$

where Φp and Θq are seasonal polynomial of Q and P, respectively. Ranking of multiplicative-seasonal ARIMA models is in form of (p, d, q) (P, D, Q)*. In this study, we attempt to model the flow rate using time series multiplicative-seasonal ARIMA model.

The Criteria for Assessing Model Performance and Error

In this study, two numerical criteria are used to assess performance and error of each grid as well as its ability in accurate prediction.

Root mean square error (RMSE) is calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}}$$

In the above equation, $\hat{y}_i$ is observational results, $\hat{y}_i$ is computational results, and N is the total number of observations. RMSE shows the difference between observed and calculated values. The lowest RMSE value shows the highest accuracy of prediction.

R^2 indicates efficiency rate of the network and is presented as follows:

$$R^2 = 1 - \frac{\sum(y_i \dots}{\sum(y_i \dots}$$

The most optimal solution for the model will be created when RMSE moves toward zero and R2 moves toward one. In time series, in addition to the above steps, Ressidual Analysis and Akaike Information Criterion should also be considered.

RESULTS AND DISCUSSION

According to the research objectives, the targeted data for the simulation of the flow rate of the area have been sorted out on a monthly basis and have been normalized using the following formula:

$$x_{norm} = \frac{x - \dots}{x_{max} - \dots}$$

Where X is non-normalized data, X norm is normalized data, X max is maximum input data, and X min is minimum input data.
Then, various models are utilized to simulate of flow rate of the basin under study by the use of adaptive neural fuzzy inference systems. Then, the most appropriate model (using the evaluation criteria) is selected from the different models to simulation the flow rate. Next, modeling of the flow rate of the basin under study is performed using time series models (ARMA-ARIMA). Ultimately, the best model of flow rate of the basin under study is selected by comparing models of neuro-fuzzy inference systems and time series models (ARMA-ARIMA).

Patterns of Flow Rate of the Basin under Study

Model 1: precipitation (P_t), temperature (T_t), evaporation (E_t) and flow rate of the drainage basin under study (Q_t)
Model 2: precipitation (P_t), temperature (T_t), evaporation (E_t) and flow rate of the drainage basin under study (Q_{t-1}, Q_t)
Model 3: precipitation (P_{t-1}, P_t), temperature (T_t), evaporation (E_t) and flow rate of the drainage basin under study (Q_{t-1}, Q_t)
Model 4: precipitation (P_t), temperature (T_t), evaporation (E_{t-1}, E_t) and flow rate of the drainage basin under study (Q_{t-1}, Q_t)
Model 5: precipitation (P_t), temperature (T_{t-1}, T_t), evaporation (E_t) and flow rate of the drainage basin under study (Q_{t-1}, Q_t)
Model 6: precipitation (P_t), temperature (T_t), evaporation (E_t) and flow rate of the drainage basin under study (Q_{t-2}, Q_{t-1}, Q_t)

Modeling Flow Rate Using Adaptive Neural Fuzzy Inference System

In this stage, modeling the flow rate of the basin under study is performed by the use of adaptive neural fuzzy inference system using Matlab 2013a software package. 80% of data is used for network training and the remaining 20% is used for network testing.

Table 1: The Results Obtained from Modeling the Flow Rate by Neural Fuzzy Model

Input Pattern	Membership Function	R^2	RMSE
1	**gaussmf**	**0.959**	**0.060**
2	**gaussmf**	**0.931**	**0.075**
3	**gauss2mf**	**0.943**	**0.106**
4	**gaussmf**	**0.927**	**0.071**
5	**gaussmf**	**0.937**	**0.092**
6	**dsigmf**	**0.946**	**0.080**

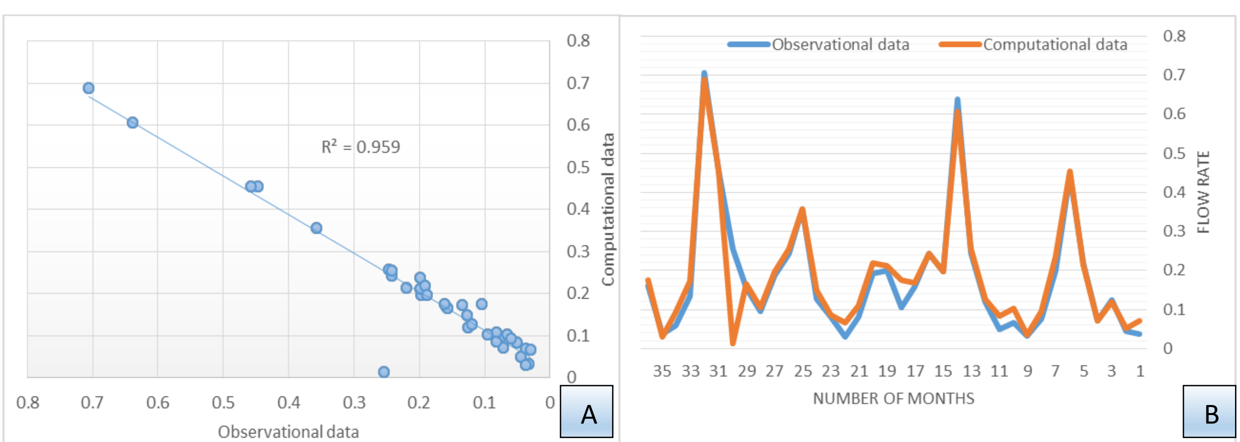

Figure 3: A- The Results of Neural Fuzzy Model for Flow Rate; B- Diagram of Comparing the Computational Data of the Selected Model with the Observational Data of the Flow Rate

Modeling Using Stochastic Models

Modeling Using Stochastic ARMA Models

In this stage, modeling the flow rate of the basin under study is performed by the use of stochastic models (ARMA) using ITSM software package. 80% of data is used for network training and the remaining 20% is used for network testing.

Table 2: The Results Obtained from Modeling of Flow Rate by Stochastic Model (ARMA)

S. No.	Flow Rate	Best AR	Best MA	Model	R^2	RMSE
1	Flow rate	1	11	ARMA (1,11)	0.765	0.270

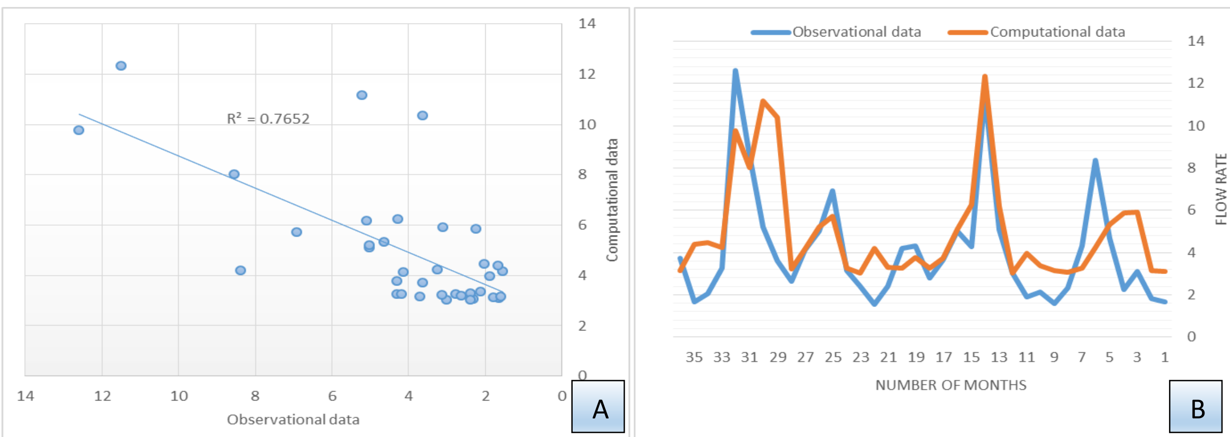

Figure 4: A- The Results of ARMA Model for Flow Rate; B- Diagram of Comparing the Computational Data of ARMA Model with the Observational Data of the Flow Rate

Modeling Using Stochastic Arima Models

In this stage, modeling the flow rate of the basin under study is erformed by the use of stochastic models (ARIMA) using Statgraphics software package. 80% of data is used for network training and the remaining 20% is used for network testing.

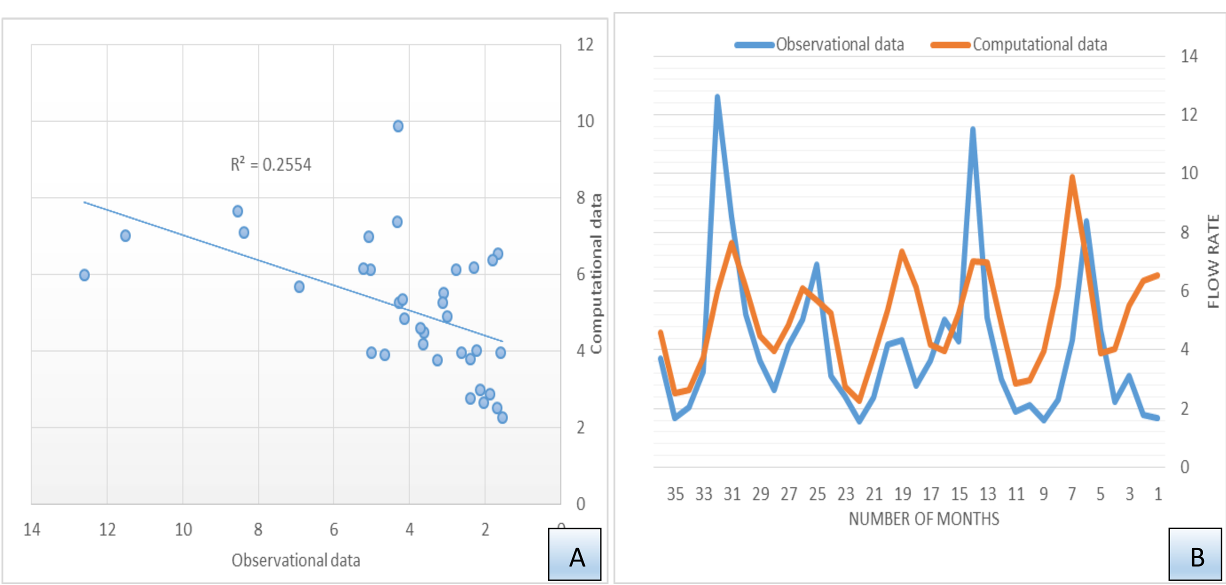

Figure 5: A- The Results of ARIMA Model for Flow Rate; B- Diagram of Comparing the Computational Data of ARIMA Model with the Observational Data of the Flow Rate

Table 3: The Results Obtained from Modeling of Flow Rate by Stochastic Model (ARIMA)

S. No.	Flow Rate	Model	R^2	RMSE	AIC
1	Flow Rate	ARIMA(1,0,0)x(2,1,2)12	0.2554	0.230375	1.721

Evaluation of Models of Neural Fuzzy System, and Stochastic Models

Table 4: Comparison of Models of Neural Fuzzy System and Stochastic Models

S. NO.	Flow Rate	Anfis Model		ARMA Model		ARIMA Model		The Best Model
		R^2	RMSE	R^2	RMSE	R^2	RMSE	
1	Flow Rate	0.959	0.0603	0.7652	0.269535	0.2554	0.230375	Anfis model

Conclusion

In this study, the performances of adaptive neural fuzzy system models as well as stochastic models were evaluated in order to simulate flow rate of Shafaroud river drainage basin in Guilan province. The results showed good performance and higher ability of neural fuzzy models in comparison with stochastic models with respect to the simulation of flow rate.

REFERENCES

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000). Artificial neural networks in hydrology II: Hydrologic applications. *Journal of Hydrologic Engineering* **5**(2) 124–137.

Chau KW, Wu CL and Li YS (2005). Comparison of several flood forecasting models in Yangtze River. *Journal of Hydrologic Engineering* **10**(6) 485- 491.

Firat M (2007). Artificial intelligence techniques for river flow forecasting in the Seyhan River catchment, Turkey. *Hydrology and Earth System Sciences Discussions, European Geosciences Union* **4** 1369-1406.

Gorr WL, Nagin D and Szczypula J (1994). Comparative study of artificial neural network and statistical models for predicting student grade point averages. *International Journal of Forecasting* **10** 17-34.

Lewis PAW and Ray BK (2002). Nonlinear modeling of periodic threshold auto regressions using TSMARS. *Journal of Time Series Analysis* **23**(4) 459–471.

Maier HR and Dandy GC (1996). The use of artificial neural networks for the prediction of water quality parameters. *Water Resources Research* **32**(4) 1013–1022.

Mishra AK, Desai VR and Singh VP (2007). Drought forecasting using a hybrid stochastic and neural network model. *Journal of Hydrologic Engineering* **12**(6) 626-638.

Nayak PC, Sudheer KP, Rangan DM and Ramasastrin KS (2004). A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrologic Engineering* **291** 52-66.

Yurekli K and Kurunc A (2005). Performances of stochastic approaches in generating low stream flow data for drought analysis. *Journal of Spatial Hydrology* **5**(1) 20-31