Journal of Engineering and Applied Sciences 11 (Special Issue 2): 3185-3192, 2016
ISSN: 1816-949X
© Medwell Journals, 2016

Use of ANFIS for Rainfall-Runoff Predictions (Case Study: Chehel-Chai Watershed, Golestan Province, Iran)

[1]Seyed Hamed Shakib, [2]Hamid Shoja Rastegari and [1]Ali Rezvani Mahmouei
[1]Department of Civil Engineering, Bozorgmehr University, Qaenat, Iran
[2]Young Elite Sponsors Institution, Tehran, Iran

Abstract: Rainfall-Runoff modeling is considered as one of the major hydrologic processes and it is essential for water resources management. In this study, runoff by the corresponding precipitation was forecasted using the ANFIS (Adaptive Neuro-Fuzzy Inference System) in Chehel-Chai basin, Golestan province, Iran. Given that the aim of this study is to predict the runoff using its corresponding precipitation and also the stations had a proper spatial and height distribution toward the basin, this basin was used in order to obtain the daily rainfall by Digital Elevation Model (DEM). Using topographic maps, the DEM Model was prepared in GIS Software and using the model, the physiographic parameters of the basin was accurately calculated. Using the DEM, digital model of the precipitation was prepared in SPSS Software. In these systems three different input combination including the same day rainfall, same day and the day before rainfall, same day and the day before and two days before rainfall were used. Due to lack of other hydrological data in this basin stations, only the precipitation and its other combinations were selected as inputs so the basin runoffs could be forecasted using a single input in ANFIS. Therefore in the ANFIS network, different types of functions were used; Gaussian, triangular and bell. The results show that the best simulation occurred in a triangular function with a combination of input with two delays. In the final step the rainfall-runoff data were classified into 10 deciles. The results show that in most deciles the model has a high performance. But the ANFIS Model was not effective in forecasting the runoff in high deciles due to the non-repetition of the model parameters of the extent year in the network education and calibration period.

Key words: Runoff forecasting, ANFIS, digital elevation model, GIS, SPSS

INTRODUCTION

The Rainfall-Runoff (R-R) process is highly non-linear and is affected by a variety of factors including rainfall characteristics, watershed morphology, soil moisture, etc., to-date many methods and approaches have been introduced to model the R-R relationship. These methods can be categorized into two main groups; physically-based models and system theoretic models. Although, physically based models help us in understanding the physics of hydrological processes, they require more data and sophisticated mathematical tools and significant user expertise. On the other hand, system theoretic models apply a different approach to identify a direct mapping between rainfall and runoff, without the need for a detailed consideration of the physical processes. Linear time series models like Auto Regressive Moving Average with Exogenous Inputs (ARMAX) and other linear and non-linear regression models, Artificial Neural Networks (ANN) and Neuro-Fuzzy Systems (NFS) are examples of this group. These kinds of models are fast and their results are often comparable with physicallybased models; however, they do not give us any information about the physics of the problem.

Over the past decade, Artificial Intelligence (AI) techniques such as Artificial Neural Networks (ANN) and fuzzy logic (Zadeh, 1965) algorithms which mimic human perception, learning and reasoning to solve complex problems, have increasingly become popular in rainfall runoff modeling. The ANN technique has been used for rainfall-runoff modeling (Shamseldin, 1997), flood forecasting (Dawson and Wilby, 1998; Rajurkar *et al*., 2002) and monthly river flow prediction (Tokar and Markus, 2000). Firat and Gungor (2007) used the ANFIS model in order to estimate the daily discharge of the Great Menderes river in the East of Turkey. Their results indicated that the ANFIS Model had a great accuracy and reliability in order to estimate the river's flow. Their results showed that the both ANN and ANFIS Models predicted the daily observational discharge of the Karaset basin properly. The performance of the two models ameliorated

Corresponding Author: Seyed Hamed Shakib, Department of Civil Engineering, Bozorgmehr University, Qaenat, Iran

when the number of the inputs increased. However, the ANFIS method showed a better performance than the ANN Model especially for prediction of the peak discharge.

Hundecha *et al*. (2001) used the fuzzy logic method in rainfall-runoff modeling. They determined the runoff from the rainfall in Neckar River basin, in southwest of Germany. In their research a conceptual modular and semi-distributed model was used which was named Hydrological Byrans Vatten Balansa Vdelning (HBV) and the best input data were presented for modeling rainfall-runoff process in studied watershed.

Nayak *et al*. (2004) utilized the Adaptive Neuro Fuzzy Inference System (ANFIS) to model the river flow rate at the Baitarani River in India. The used model in their research had good performance on base of various statistical indices. Tayfur and Singh (2006) used ANN and fuzzy logic for predicting event based rainfall-runoff and tested these models against the Kinematic Wave Approximation (KWA). The results provided insights into the adequacy of ANN and Fuzzy Logic (FL) methods as well as their competitiveness against the KWA for simulating event-based rainfall-runoff processes.

Fuzzy logic models have been applied to simulate river discharges (Hundecha *et al*., 2001), predict runoff (Xiong *et al*., 2001; Sen, 2006; Sen and Altunkaynak, 2006; Alexandra and Asaad, 2006; Lohani *et al*., 2011) and forecast water supply from snow melt (Mahabir *et al*., 2003). Hybrid systems such as neuro-fuzzy systems combine ANN and fuzzy systems to compensate for weaknesses in individual systems. Aqil *et al*. (2007) used ANN to optimize parameters of Takagi-Sugeno rule-base of a neuro-fuzzy model and compared the performance of it with two other ANN Models to predict the flow of the Cilalawi River in Indonesia. The neuro-fuzzy Model has outperformed the other ANN models. A fuzzy expert model for estimating the index water yield and index flow (the index flow is the median flow for the lowest flow month of the flow regime) of ungauged streams in Michigan was also reported as the most robust method among other models tested; multiple linear regression, fuzzy regression and adaptive neurofuzzy inference models (Hamaamin *et al*., 2013).

Background investigations on the simulation of the runoff-precipitation process using the ANFIS show that the most of these studies have attempted to predict the long-term runoff. In those studies which simulated the rainfall-runoff process in short terms, the simulation was based on shower. The aim of this study was to investigate the possibility of the runoff simulation without knowing the other hydraulic data from the basin other than the precipitation by the ANFIS network. Therefore, in this study the calculation of the average regional rainfall will be first discussed in the GIS and SPSS Software using the DEM Model. Then modeling of rainfall-runoff with different application of diverse precipitation combination was conducted as different input and member functions in ANFIS using the MATLAB Software.

MATERIALS AND METHODS

Case study (Chehel Chai basin): Chehel Chai basin with an area of 25683/12 ha is located between the longitude of 55°23-55°38 and the latitude of 36°59-37°13. This basin is a subdivision of the great Gorganrood basin that is located within the city boundary of Minoodasht in Golestan province. It has a minimum and maximum elevation of 190 and 2570 m above the free surface sea level.

Figure 1 shows the location of Chehel-Chai basin. The Chehel Chai basin was divided into three subbasins in the ArcGIS environment according to the basin topography and drainage network. In next Step the physiographic characteristics such as weighted average slope, weighted average elevation, area, perimeter and etc. were determined with respect to sub-basin separation. This date is shown in Table 1. In the Chehel-Chai basin, the lazoreh hydrometric station was used as a base station around the basin outlet in order to calibrate and verify the model. The information on the stations around and inside of the Chehel-Chai basin is presented in Table 2 and the positions of these stations are shown in Fig. 2.

Runoff in the outlet hydrometric station of the basin (lazoreh) shows the produced runoff in the entire basin. Therefore In the first step, average precipitation of basin level should be calculated for its corresponding rainfall. Recorded numbers in each weather station show the spot precipitation amount of the same station, exclusively. In order to determine the rainfall amount of the basin level, we should take average of the recorded rainfall in the stations based on the inside and outside of the limited area. Different interpolation methods such as Kriging, IDW (Inverse Distance Weight), etc. or the relationship between precipitation and geographical characteristics such as length width and height can be used to determine the regional average precipitation. In this study, the digital elevation model or DEM was used to calculate the regional average precipitation. Given that the correlation coefficient between rainfall and elevation and the F test in 5% level are meaningful and also the remainders in the most years which were studied are normal, the digital elevation model was used to calculate the regional average precipitation. It is noteworthy that the SPSS Software was used for statistical analysis. The following

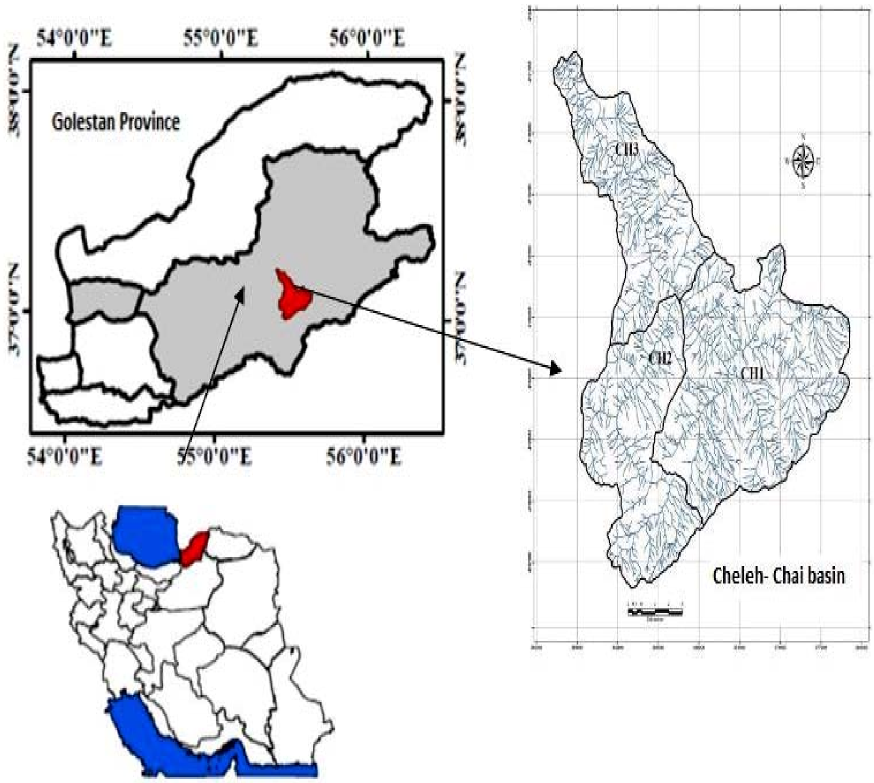

Fig. 1: The location of Chehel-Chai basin in Golestan province

Table 1: Chehel-Chai sub-basins physiographic characteristics

Sub-basin	Area	Perimeter	Weighted average slope (%)	Weighted average height (m)
CH1	118.82	53.89	46.89	1388.80
CH2	71.26	45.99	42.88	1327.31
CH3	66.75	46.19	47.06	713.56
Total	256.83	95.24	45.82	951.14

Table 2: Characteristics of the chehel chai basin stations

Station	X (m)	Y (m)	Z (m)	Station type	Start year period
Chehl-Chai	358258	4120965	190	Hydrometer	1971
Lazoreh	358269	4120983	155	Hydrometer	1978
Jangaldeh	353505	4114636	180	Hydrometer	1971
Gholitapeh	359548	4121932	250	Hydrometer	1975
Narab	374176	4097435	1500	Rain gauge	1980
Farsian	375996	4121020	900	Normal rain gauge	1975
Galikesh	361086	4125346	250	Stable rain gauge	1971

steps were taken in order to obtain the regional average precipitation by the digital elevation model using the Arc GIS Software:

- Creating the input files of the Arc GIS Software
- Obtaining the relation between rainfall and elevation using the regression equations
- Obtaining a digital map of the study area for each years
- Analyzing of the Digital Elevation Model (DEM)
- Analyzing and drawing of the spatial variations of rainfall in the region for each year
- Determination of the rainfall amount in the study region for each years

Due to the impact of cellular network size on the accuracy of digital elevation model, the cellular network size was recognized after the investigation of contour lines distance and therefore the DEM of the basin with a spatial resolution of 10×10 m was prepared Fig. 3 shows a map of Chehel-Chai basin DEM. According to the taken steps above, the rainfall-runoff data were made for modeling using the ANFIS system.

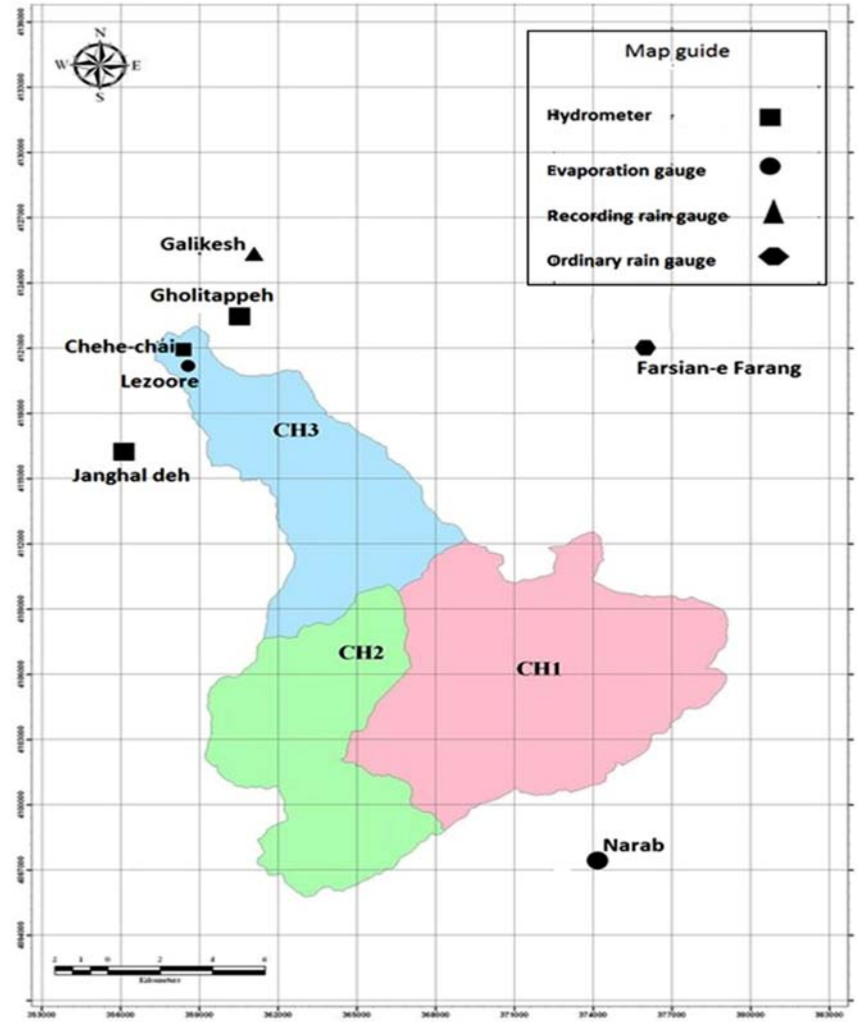

Fig. 2: The location of Chehel-Chai basin stations

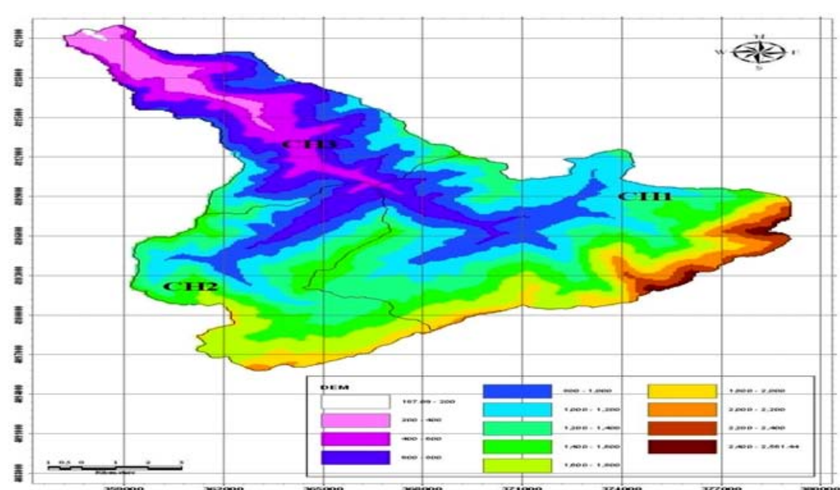

Fig. 3: Chehel-Chai basin DEM map

RESULTS AND DISCUSSION

Simulation: First, the structure of specific model parameters (proportional to the input, the input membership functions and rules and output membership functions and output variable) is assumed. Then a series of input/output which is usable by phase inference/neural adaptive is collected. Then using the phase inference/neural adaptive, the model of the phase inference system is taught using the existing data then we can modify the parameters of membership functions according to the selected error so the data from the model will get closer to the actual value. In most cases the data is collected with some errors and the used data in training can't be a representative of the whole combination of the data which will be provided to the model during the utilization period. Model validation is a process during which the input vectors of input/output data section that are not used for the training of the phase inference are used as input data in the developed model so we can gather information from the prepared phase inference system in order to predict the output values of the input corresponding data. This task is done using the test data collection. The validation of the phase inference-neural adaptive can be determined using another sample of data collection (control data). The model potential control in the more fitting-field begins on the training data. As a general rule, the model error for the control data collection reduces from the beginning of the training process until the time when the more fitting-process starts. Therefore the data is normalized between the values of zero and one using the Eq. 1 and is prepared to enter the models:

$$X_1 = (X_1 - X_{min})/(X_{max} - X_{min}) \quad (1)$$

In this Eq. 1 X and Y are the maximum and minimum values of statistical series respectively and X_i is the value of each available data in the statistical series. After the initial investigations, the available parameters in an input vector were organized with a title of each incident rainfall and an output vector (Discharge). The used length of statistical period was 26 year (9497 day) for rainfall-runoff which 19 year of this period (1990-2008) were for education and 7 year (2009-2015) were for experimentation of these two models. The statistical characteristics of the daily rainfall and runoff data of the ChehelChai basin is provided in Table 3.

The coding in Matlab Software was used In order to simulate in ANFIS system. The model is one of the models which treat the system as a black box and in the training stage set its parameters in a way that they can provide similar outputs, given the various inputs using the input and output data. In fact the control data cause the reduction of the model's more fitting. If there is a significant difference between control and training data, the non-performance of the control data collection for the purpose of the model validation will be decreased by gaining the control error during the training period. For this purpose, various combinations of input data including the same-day rainfall (R), same day and the day before rainfall (R and R-1) and the same day and the day before and two days before(R and R-1 and R2) were used. Each of these input combinations were evaluated by triangular membership function, Gaussian type 1-2 and Gaussian bell. Also for grouping of the input data, the cluster separation was used. In this study, the sharing operator (AND), the assembly Operator product (OR) and the maximum method were used in order to teach ANFIS. The product method (Prod) and the Maximum method (Maximum) were used respectively for inferring and gathering of information and for the defuzzification method, weighted average method was used. To compare the results of the models, the three criteria of correlation coefficient, efficiency coefficient and root mean square error were used (Eq. 2-4):

$$R^2 = \frac{\sum(O_i - \bar{O})(p_i - \bar{p})}{\sqrt{\sum(o_i - \bar{o})^2 \sum(p_i - \bar{p})}} \quad (2)$$

$$RMSE = \left(\frac{\sum_{i=1}^{n}(p_i - o_i)^2}{n}\right)^{1/2} \quad (3)$$

Table 3: Statistical data of the daily rainfall and runoff of chehel-chai basin

Variables	Data series	Data numbers	Average	Minimum amount	Maximum amount	SD
Precipitation (mm)	Total	6940	1.44	0	105.37	4.12
	Exam	2557	1.69	0	46.04	4.79
	Training	9497	1.51	0	105.37	4.21
Runoff (M3S-1)	Total	6940	1.94	0	248	6.64
	Exam	2557	2.12	0	267	8.74
	Training	9497	1.99	0	267	7.27

Table 4: Results of ANFIS network with membership functions and different combinations of input in chehel-chai basin

Model input	Rules number	Membership function	R^2 (%)	R^2	RMSE
Rt	2	Triangular	83.6	6.87	0.8340
		Bell	82.42	6.90	0.8242
		Gaussian	81.6	6.89	0.8160
		Gaussian type 2	82.08	6.90	0.8208
Rt Rt-1	4	Triangular	87.6	6.70	0.8760
		Bell	87.42	6.74	0.8742
		Gaussian	86.7	6.73	0.8670
		Gaussian type 2	87.07	6.71	0.8707
Rt	8	Triangular	94.45	6.62	0.9445
Rt-1		Bell	93.08	6.65	0.9308
Rt-2		Gaussian	91.98	6.64	0.9198
		Gaussian type 2	93.22	6.60	0.9322

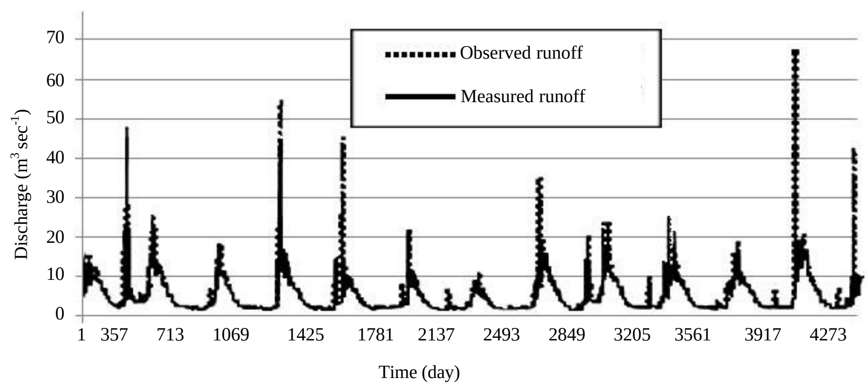

Fig. 4: Comparison of the observed and predicted runoff with combination of Rt input with ANFIS model

$$r^2 = 1 - \left(\frac{\sum_{i=1}^{n} (o_i - p_i)^2}{\sum_{i=1}^{n} (o_i - \bar{o})^2} \right)^{1/2} \quad (4)$$

In these Eq, O and P are the observed and simulated values respectively. O and P are average of the observed and simulated values respectively and N is the number of the samples.

The diverse membership function can be used in the phase-neural system which might be effective in the final result. So the same input combinations in various triangular membership functions, Gaussian, Gaussian type 2 and bell were used. The results of this simulation and statistical profile related to them are provided in Table 4. As can be seen in Table 4, by changing the membership function, the results don't vary very much but the triangular function is better than other membership functions. Also by increasing the number of inputs in every four functions, the simulation accuracy increases in a way that the best simulation in triangular function occurred with combining the input with two delays. These results are consistent with Kurtulus and Razak (2010). Adaptive phase-neural network, triangular function with combination of two delay inputs (Rt, Rt-1, Rt-2) has provided the best simulation. In this system, the different rainfall combinations of the ChehelChai basin (Rt, Rt-1, Rt, Rt-1, Rt, Rt-2) were used as input. With an increase in number of the inputs, the simulation accuracy improved. So that the, most accuracy in the model is provided by combining the input with two delays. This reflects the impact of high runoff from the Cheleh-Chai basin from the rainfall of the preceding 2 days. Figure 4-6 also reflect the ANFIS model for the estimated amounts of the runoff due to different rainfall combinations in the Chehel-Chai basin associated with the measured values.

Then, in order to clarify the cause of the network inability to determine the high level data, we proceeded the determination of the data statistic percentile in the statistical deciles. In this way the distance between the minimum and maximum value in any given data category is divided into ten equal parts and the percentage of the total data in each deciles is shown in Table 5. The results of Table 5 show that the number of upper limit data is very low compared to the total data and this could be the most important network incompetency factor for the proper simulation of the upper limit data.

The results from Table 5 show that >90% of the data is located in the first two deciles and in the Lazoreh station, <1% of the total data is located in the last

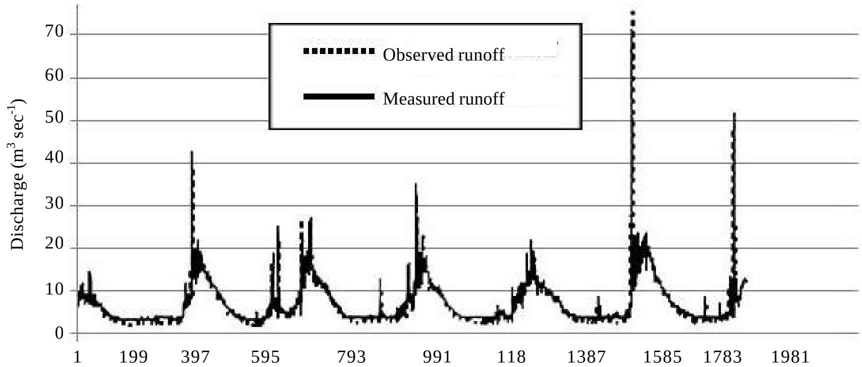

Fig. 5: Comparison of the observed and predicted runoff with combination of R_t input, R_{t-1} with ANFIS model

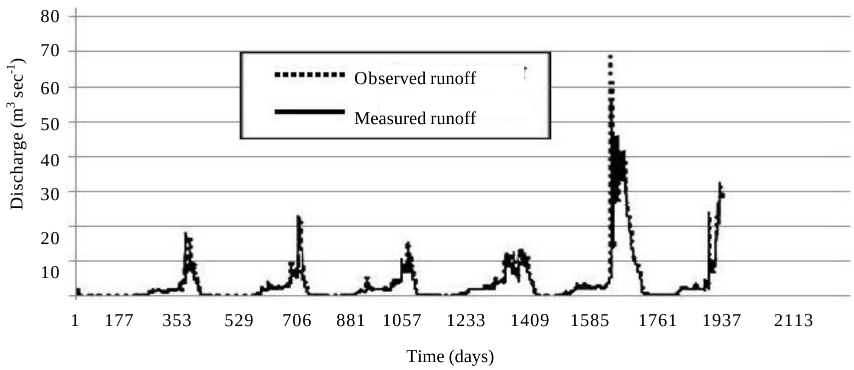

Fig. 6: Comparison of the observed and predicted runoff with combination of R_t input, R_{t-1}, R_{t-2} with ANFIS model

4 deciles (that the maximum numbers and upper limit discharge should be located in those deciles). This number is <9 data numbers among the 9497 data numbers. If all this data in a form of a training group is given to the network for training, Due to the highly complex relationship between input and output parameters, one can not expect too much of the phase-neural system for the proper training and a correct response provision. So, it is natural that the provided data by the model about the upper limit discharge differ from the observational data.

CONCLUSION

Obviously, climate and rainfall are great non-linear phenomena in nature, known as butterfly effect. Using the phase system for predicting the runoff can be considered the most appropriate method due to the high results accuracy. Therefore, this method was used in order to create the total precipitation data of the basin in the GIS. So for the precipitation of any event in the whole basin, first the calculation of the average rainfall in the basin must be done. The DEM method was used in order to obtain the average amount of rainfall for the entire basin. Then the average rainfall data was selected as input and the runoff of the Lazoreh hydrometric station was selected as output in ANFIS model. For this purpose, various combinations of data input including the same day Rainfall (R), the same day and the day before rainfall (R and R-1) and the same day, the day before the same day and two days before the same day (R and R-1 and R-2) were used. Each of these input combinations with triangular membership functions, Gaussian type 1 and 2, Gaussian and bell were evaluated. Also the cluster separation was used in order to group the input data. With an increase in number of data the simulation accuracy improved. The highest accuracy in the model in a triangular function was provided combining the input with two delays. This reflects the high impact rate of the Chehel-Chai basin runoff from the rainfall of its two preceding days. As it is obvious from the results, the phase-neural system is not able to provide the proper answer in the prediction of the runoff before the addition of the previous rainfall data to the data collection. The data survey showed that the relationship between rainfall and runoff in the study area is a complex relation. This means that sometimes a certain amount of rainfall has caused a considerable amount of runoff but in other cases the same amount of precipitation has caused much lower volume of runoff or has not made a considerable change in the base discharge of the river. Also, we determined the data statistical percentage in statistical deciles in order to clarify the system's incompetency in determination of the upper limit data. About >90% of the whole data is in the first two deciles and >1% of the whole data is in the last fourth deciles. Therefore, the ANFIS model was not effective in prediction of the runoff in high deciles of rainfall and this is because of the little rainfall data for training the model and it is not because of the model itself 0.333.

REFERENCES

Alexandra, p. and Y.J. Asaad, 2006. Development of rainfall-runoff models using takagi-sugeno fuzzy inference system. J. Hydrol., 329: 154-173.

Aqil, M., I. Kita, A. Yano and S. Nishiyama, 2007. A comparative study of artificial neural networks and neuro-fuzzy in continuous modeling of the daily and hourly behaviour of runoff. J. Hydrol., 337: 22-34.

Dawson, C.W. and R. Wilby, 1998. An artificial neural network approach to rainfall-runoff modeling. J. Hydrol. Sci., 43: 47-66.

Firat, M. and M. Gungora, 2007. River flow estimation using adaptive neuro fuzzy inference system. Math. Comput. Simul., 75: 87-96.

Hamaamin, Y.A., A.P. Nejadhashemi and M.D. Einheuser, 2013. Application of fuzzy logic techniques in estimating the regional index flow for Michigan. Trans. ASABE., 56: 103-115.

Hundecha, Y., A. Bárdossy and H. Theisen, 2001. Development of a fuzzy logic-based rainfall-runoff model. Hydrol Sci. J., 46: 363-375.

Kurtulus, B. and M. Razack, 2010. Modeling daily discharge responses of a large karstic aquifer using soft computing methods: Artificial neural network and neuro-fuzzy. J. Hydrol., 381: 101-111.

Lohani, A.K., N.K. Goel and K.K.S. Bhatia, 2011. Comparative study of neural network, fuzzy logic and linear transfer function techniques in daily rainfall-runoff modelling under different input domains. Hydrol. Proces., 25: 175-193.

Mahabir, C., F.E. Hicks and R. Fayek, 2003. Application of fuzzy logic to forecast seasonal runoff. Hydrol. Process, 17: 3749-3762.

Nayak, P.C., K.P. Sudheer, D.M. Rangan and K.S. Ramasastri, 2004. A neuro-fuzzy computing technique for modeling hydrological time series. J. Hydrol., 291: 52-66.

Rajurkar, M.P., U.C. Kothyari and U.C. Chaube, 2002. Artificial neural networks for daily rainfall-runoff modelling. Hydrol. Sci. J., 47: 865-877.

Sen, Z. and A. Altunkaynak, 2006. A comparative fuzzy logic approach to runoff coefficient and runoff estimation. Hydrol. Processes, 20: 1993-2009.

Shamseldin, A.Y., 1997. Application of a neural network technique to rainfall-runoff modelling. J. Hydrol., 199: 272-294.

Tayfur, G. and V.P. Singh, 2006. ANN and fuzzy logic models for simulating event-based rainfall-runoff. J. Hydrol. Eng., 132: 1321-1330.

Tokar, A.S. and M. Markus, 2000. Precipitation-runoff modeling using artificial neural networks and concettual models. J. Hydrol. Eng., 5: 156-161.

Xiong, L., A. Shamseldin and K.O ' Connor, 2001. A non-linear combination of the forecasts of rainfall-runoff models by the first-order Takagi-Sugeno fuzzy system. J. Hydrol., 245: 196-217.

Zadeh, L.A., 1965. Fuzzy sets. Inform. Control, 8: 338-353.