I.J.S.N., VOL.7 (4) 2016: 714-723 ISSN 2229 - 6441

INTERNATIONAL JOURNAL OF SCIENCE AND NATURE

www.scienceandnature.org

A COMPARATIVE STUDY OF ADAPTIVE NEURO FUZZY INFERENCE SYSTEM (ANFIS) AND MULTIPLE LINEAR REGRESSION (MLR) FOR RAINFALL-RUNOFF MODELLING

[*1]Vijay Kumar Singh, [2]Pravendra Kumar , [3]Bhaskar Pratap Singh and [4]Anurag Malik
[1, 2, 4] Department of soil and water conservation engineering, Gobind Ballabh Pant University of Agriculture and Technology, Pantnagar, Uttarakhand, INDIA
[3]Department of farm engineering, Institute of Agricultural Sciences, BHU, Varanasi
[*]Corresponding author email: vijaysinghswce@gmail.com

ABSTRACT
Runoff prediction has an important role in water management and flood prediction. Floods are one of the most serious natural disaster and present major social concerns. The effective flood management is always of great apprehension in the field of hydrology and water resources engineering. In the present study, two different techniques namely adaptive neuro-fuzzy inference system (ANFIS) and multiple linear regression (MLR) has been employed, to estimate daily runoff for Arpa river, Bilaspur, Chhattisgarh, India. The data of monsoon season (June 1st to September 30th) of five years (2001-2005) were used for training of models and data of two years (2006-2007) were used for testing of developed models. ANFIS model trained used to two different type membership function (MFs) viz. Gaussian and Generalized bell, and conventional techniques was used multiple linear regression (MLR). Also Gamma test (GT) was used to finding the reliable input combination of input variables. The performance of the models were evaluated qualitatively by visual observation and quantitatively using different performance indices viz. root mean square error (RMSE), correlation coefficient (r), coefficient of efficiency (CE) and t-Test. It was found that the performance of the ANFIS model is better than the MLR models in estimation of daily runoff for Arpa River.

KEYWORD: adaptive neural-fuzzy inference system, multiple linear regression, Gamma test, t-Test.

INTRODUCTION

Water is the most appreciated and important gift in the natural surroundings and it must be conserved and maintained carefully for all living and non-living belongings. Due to shortage of water and increased rate of population, its turn out to carry and the optimal use of available water resources for proper planning and efficient water resources in order to forecast stream flow for purposes such as water supply, floods control, irrigation, drainage, water quality. Forewarning of floods can indeed go a long way in preventing much potential damage due to floods. The efficient use of available water resources calls for proper planning, design, operation, and management of the existing water resources using advanced technologies. A key component in the water resources planning, development, design operation, and management systems is the accurate estimation of the runoff at a local source such as a River. One of the major areas of research interest in water resources management is forecasting the future rainfall-runoff. A wide range of methods being used for runoff forecasting and the most of these methods based on the statistical analysis of historical data which has measured in the past. The process of conversion of rainfall into runoff over a catchment is very complex, highly nonlinear, and exhibits both temporal and spatial variability. The many models have developed to simulate this process. These can be regarded as empirical, black box, conceptual, and physically based distributed models. Each of these types of models has its own advantages and limitations. Although conceptual and physically-based models have the main tool for depicting hydrological variables and understanding the physical processes taking place in a system, they do have practical limitations. When data is not sufficient and getting accurate predictions is more important than conceiving the actual physics, empirical models remain a good alternative method, and can provide useful results without a costly calibration time. ANN model is a black box models with particular properties which is greatly suited to dynamic nonlinear system modeling. The artificial neural network (ANN) and adoption of the adaptive neuro-fuzzy inference system (ANFIS) techniques for rainfall-runoff modeling is adding a new dimension to the system theoretic modeling approach and it is applying in recent years, as a successful tool to solve various problems. In recent years, a great deal of work has been done in applying data driven models like multiple regressions and neural networks for water resources research. Conventional multiple linear regression (MLR) methods have been widely used in daily rainfall-runoff modelling, water level forecasting at gauging stations, real time flood forecasting and warning. Dawson and Wilby (1999) demonstrated that multi-layered perceptron has a better performance than MLR method in one step ahead river flow predicting, using past rainfalls and discharges. Bisht *et al.* (2010) presented that multilayer feed-forward ANN models are superior to MLR models in forecasting one-step ahead discharge, using past River stages and discharges. Asati and Rathore (2012) compared the ANN with autoregressive (AR) models and multiple-linear regression (MLR) for short-term flow prediction. Recently established attention in hydrology is the fuzzy-rule based approach in modeling. Neuro-fuzzy model was successfully used in the hydrological sciences during recent year. It is new improved tool and a data driven modeling approach for determining the behavior of imprecisely defined complex dynamical systems (Panchal *et al*., 2014). Several studies carried out using fuzzy logic in hydrology and water resources planning (Kermani *et al.*, 2007; Talei *et al.*, 2010 and Bisht and Jangid, 2011). An ANFIS can simulate and analyze the mapping relation between the input and output data through a learning algorithm to optimize the parameters of a given Fuzzy Inference System (FIS). Adaptive neuro fuzzy inference system (ANFIS) is a fuzzy mapping algorithm that is based on Tagaki-Sugeno-Kang (TSK) fuzzy inference system (Loukas, 2001 and Nayak *et al.,* 2003). The accurate global learning ability of TSK model was motivated the applications of such models in non-linear system estimation. Adaptive Network-based Fuzzy Inference System (ANFIS) is an example of TSK models where global parameter tuning is implemented by minimizing the global error of the model (Jang, 1993). ANFIS provides a method for the fuzzy modeling procedure to learn information from the data set followed by creation of the membership function parameters that the best performs the given task. The most of the previous investigations have indicated that ANFIS is an efficient tool for rainfall-runoff modeling and is widely used in different areas of water related research. In the past few decades ANNs and ANFIS methods have widely used in a wide range of engineering applications including hydrology such as for rainfall-runoff modeling, groundwater modeling and River flow forecasting (Tokar and Johnson, 1999; Xiong *et al.,* 2001; Shamseldin *et al.*, 2010 and Shrivastav *et al.*, 2014). Ghose *et al.* (2013) developed the models for prediction of runoff using non-linear multiple regression (NLMR) and adaptive neuro-fuzzy Inference system (ANFIS). Rezaeianzadeh *et al.* (2013) studied the use of artificial neural networks (ANNs), adaptive neuro-fuzzy inference system (ANFIS), multiple linear regressions (MLR) and multiple nonlinear regressions (MNLR) for forecasting maximum daily flow. There are many comparative studies and application of ANN and ANFIS in field of hydrology and water resource (Nayak et al. 2004; Bisht and Jangid, 2011; Folorunsho *et al.*, 2014 and He *et al.*, 2014). In this paper, the applicability of the ANFIS approaches for modeling runoff of river is investigated. The results are the compared with multiple linear regression (MLR) models for rainfall-runoff modeling of Arpa River.

MATERIALS & METHODS

Study Area

The daily rainfall and runoff data during the period from 2001-2007 for Arpa River were recorded from Ghatora station of Central Water Commission (CWC) and the data were obtained from Divisional office of CWC Raipur, Chhattisgarh, India. The Ghatora station of Arpa River is located in Bilaspur district of the Chhattisgarh state in India at latitude of 22 33’29.16’’ N and longitude of 82 6’41.20’’ E and having elevation of 246 m from mean sea level (MSL). The drainage area of Arpa River is approximately 3035 km^2. The location of study area is shown in Fig. 1. The seven years data set are divided into two phases, first phase is training and second is testing. The models are trained using the five years data from June 1, 2001 to September 30, 2005, and the testing of the models was done using the two years data from June 1, 2006 to September 30, 2007 for validation of developed models. The statistical parameters of rainfall and runoff data are shown in Table 1.0 which indicates that the rainfall and runoff show significant skewed distribution. The ratio between the standard deviation and mean is high.

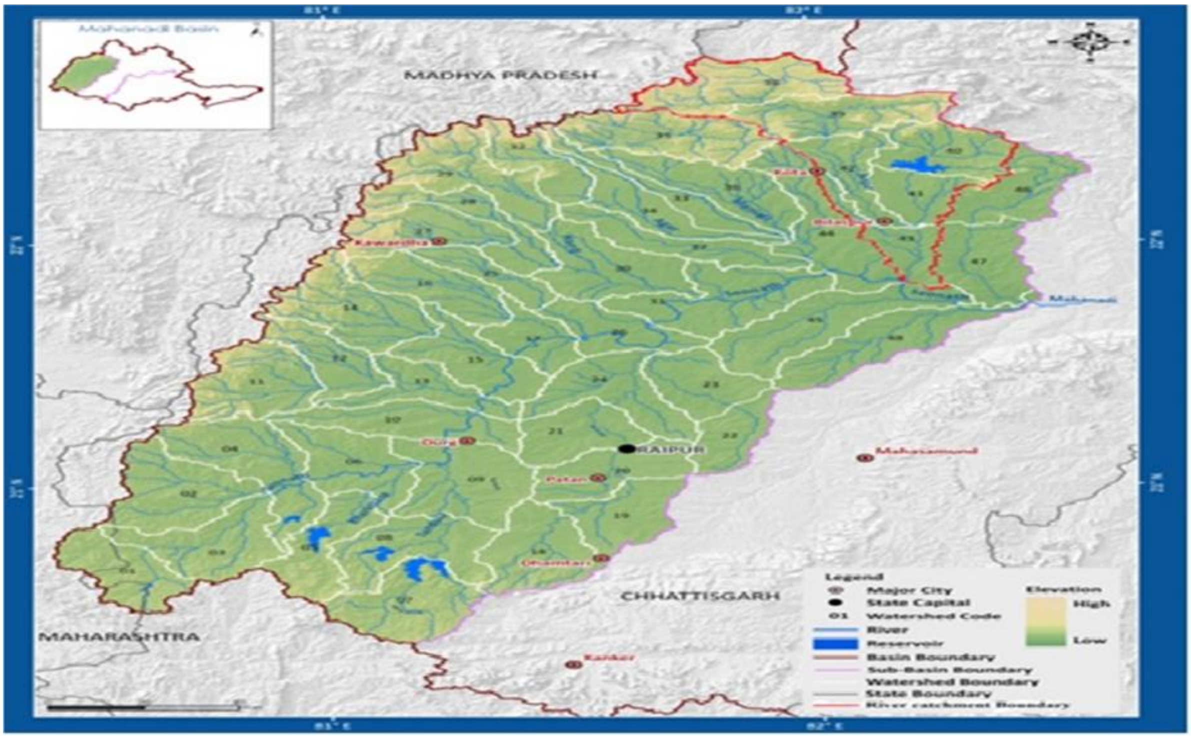


FIGURE 1.0 Location map of the Arpa catchment

TABLE 1.0 Statistical parameter of data set for training and testing at Ghatora site on Arpa River

Statistical parameters	Training data set		Testing data set	
	R_t	Q_t	R_t	Q_t
Mean	6.28	46.65	4.02	41.30
Standard deviation	14.08	81.62	9.16	42.02
Standard Error	0.57	3.30	0.59	2.69
Coefficient of Skewness	4.10	7.71	3.63	4.81
Maximum	127.10	1248.37	65.01	425.20
Minimum	0	0	0	4.01

Adaptive neuro-fuzzy inference system (ANFIS)

The adaptive neuro-fuzzy inference system (ANFIS) models, which consist of both ANN and fuzzy logic methods, were first introduced by Jang (1993). A neuro-fuzzy system integrates fuzzy inference systems (FIS) and neural networks which have the potential to capture the benefits of both methods. Fuzzy systems have the advantages of describing the fuzzy rules and being interpretable, which make it possible to represent the real world process and identify the reason of particular value in the fuzzy system output. On the other hand, fuzzy systems need expert information or directions to define fuzzy rules and tuning the parameters of fuzzy systems (e.g. membership functions parameters). By increasing the complexity of the process, developing fuzzy rules and membership functions become more difficult and sometimes impossible. In the neural networks approaches, the opposite situation can be observed. Neural networks are not able to explain the behavior of the system based on previous information, but they are trainable which gives them the ability of tuning their structures from input-output data. Considering these facts, using a hybrid model of fuzzy and neural networks eliminates these problems. However, ANFIS has more computational complexity restrictions than ANN.

ANFIS architecture

One of the most popular integrated systems is adaptive neuro-fuzzy inference system (ANFIS) which has shown promising results in modelling nonlinear time series. In ANFIS, Takagi-Sugeno type fuzzy inference system is used. The output of each rule can be a linear combination of input variables plus a constant term. The final output is the weighted average of each rule's output. Basic ANFIS architecture that has two inputs X and Y and one output Z is shown in Fig. 2.0. The rule base contains two Takagi-Sugeno if-then rules as follows:

Rule 1: If x is A_1 and y is B_1, then $f_1 = p_1x + q_1y + r_1$ …. (1)

Rule 2: If x is A_2 and y is B_2, then $f_2 = p_2x + q_2y + r_2$ …. (2)

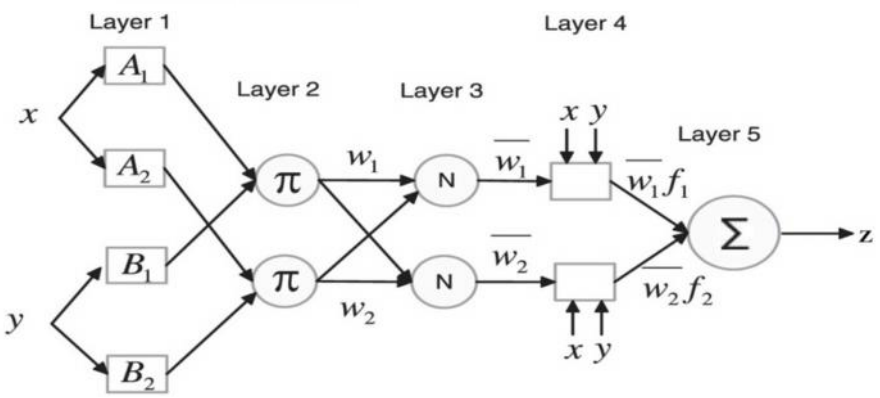


FIGURE 2.0 The ANFIS structure with two inputs

Where A_n and B_n are fuzzy sets in the antecedent; p_n, q_n and r_n are polynomial parameters of n^{t} rule (also called the consequent parameters).

The node functions in the same layer are the same as described below:

Layer 1: Every node I in this layer is a square node with a node function as:

$O_i^1(x) = \mu\, A_i(x)_i$ for i = 1, 2 (3)

$O_i^1(x) = \mu\, B_i - 2(x)_i$ (4)

Where x is the input to i^{th} node, A_i (or i−2 B) is a linguistic label (such as "small" or "large") connected with node, and O_i is the membership grade of a fuzzy set A_i such as Gaussian and Generalized bell. Based on the problem, different membership functions can be applied. For instance, if the membership function of i^{t} node is a generalized bell function, the output of i^{t} node in the first layer defines as:

$$\mu\, A_i(x) = \frac{1}{1+|(x-c)/a|^{2b}} \quad (5)$$

And the Gaussian function

$$\mu\, A_i(x) = e^{-\left(\frac{x-c_i}{a_i}\right)} \quad (6)$$

Where (a_i, b_i, c_i) are premise parameters that change the shape of the membership function.

Layer 2: Every node in this layer is a fixed node labeled as II, whose output is the product of all incoming signals**:**

$O_i^2 = W_i = \mu A_i(x)_i \, \mu B_i(x)_i \quad i = 1, 2 \quad (7)$

Layer 3: Every node in this layer is a fixed node labeled N. The i^{th} node calculates the ratio of the rule's firing strength of the sum of all rule's firing strengths:

$$O_i^3 = W_i = \frac{W_i}{(W_1 - W_2)} \quad i = 1, 2 \quad (8)$$

Layer 4: Ever node i in this layer is a square node with a node function:

$$O_i^4 = \overline{W}_i f_i = \overline{W}_i (p_i x + q_i y + r_i) \quad (9)$$

Where, $\overline{W}_i$ is a normalized firing strength from layer 3 and $\{p_i, q_i, r_i\}$ is the parameter set of this node. Parameters in this layer are referred to as "consequent parameters".

Layer 5: The single node in this layer is a fixed node labeled sigma that computes the overall output as the summation of all incoming signals:

$$O_i^5 = \sum_i \overline{W}_i f_i = \frac{\sum_i W_i f}{\sum_i W_i} \quad (10)$$

This layer is called as the output nodes in which the single node computes the overall output by summing all the incoming signals and is the last step of the ANFIS. In this way the input vector was fed through the network layer by layer.

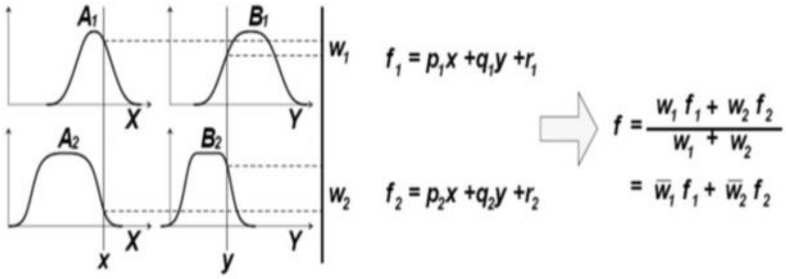


FIGURE 3.0 A two input first order Takagi-Sugeno-Kang (TSK) fuzzy model

The adaptive neuro-fuzzy inference system (ANFIS) models were trained and tested using Gaussian and Generalized bell membership functions, TSK fuzzy model, hyperbolic tangent activation function, and Delta-Bar-Delta learning algorithm. Table 2.0 explains the training variables in the adaptive neuro-fuzzy inference system (ANFIS) models. The number of membership function assigned to each input of the adaptive neuro-fuzzy inference system (ANFIS) was set to 2, 3, 4, 5 and 6 respectively. The activation function was used to gives the best result in depicting the non-linearity of the modeled natural. The learning algorithm Delta-Bar-Delta is minimizing the error in input-output data sets.

TABLE 2.0 Training variables and their assigned values for ANFIS models

Training variables	Assigned value
Membership function	Gaussian and Generalized bell
MFs per input	2 to 6
Fuzzy Model	TSK
Activation function	Hyperbolic tangent
Learning rule	Delta-Bar-Delta
Epoch	1000
Training threshold	0.001

Activation functions

The most commonly used activation functions in hydrology for the best applications are Sigmoid and hyperbolic tangent. In the present study, hyperbolic tangent activation function is used. The output range of hyperbolic tangent function is bounded into the range of -1 and 1, for inputs, which is considered as the desirable characteristics of this function.

The hyperbolic tangent activation function is mathematically expressed as:

$$Tan(x) = \frac{sin(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad \ldots (11)$$

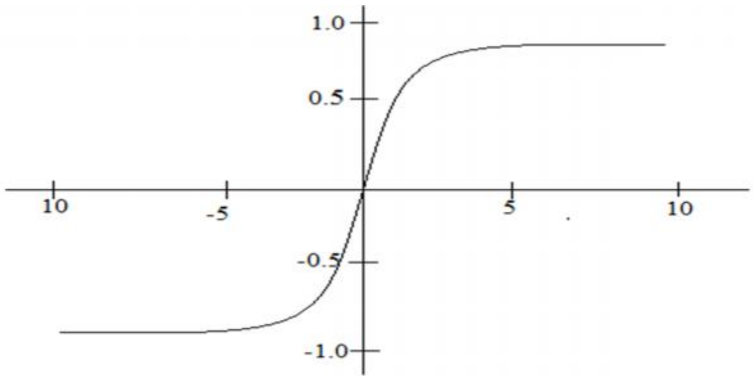


FIGURE 4.0 Hyperbolic tangent function

This function is defined as the ratio between hyperbolic sine and cosine, function as ratio of half difference and half-sum of exponential function in point x and –x, as shown in Fig 4.0. The hyperbolic tangent function is similar to sigmoid function with its outputs between -1 to 1.

Membership functions (MFs)

A fuzzy set is completely characterized by its membership function. Since most fuzzy sets in use have a universe of discourse consisting of the real line, it would be impractical to list all the pairs important in a membership function. The most common membership functions used in fuzzy sets is Gaussian and generalized bell membership functions (MFs). Because of their smoothness and concise notation, Gaussian and generalized bell membership functions (MFs) are becoming increasingly popular for specifying fuzzy sets. Gaussian function is well known in probability and statistics, and they possess useful properties such as invariance under multiplication (the product of two Gaussians is a Gaussian with a scaling factor) and Fourier transform (the Fourier transform of a Gaussian is still a Gaussian). The generalized bell MF has one more parameter than the Gaussian MF, so it has one more degree of freedom to adjust the steepness at the crossover points.

(a) Gaussian membership functions (MFs)

A Gaussian MFs is specified by two parameters;

$$gaussi(x, c, \sigma) = e^{-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2} \quad \text{…. (12)}$$

Where,

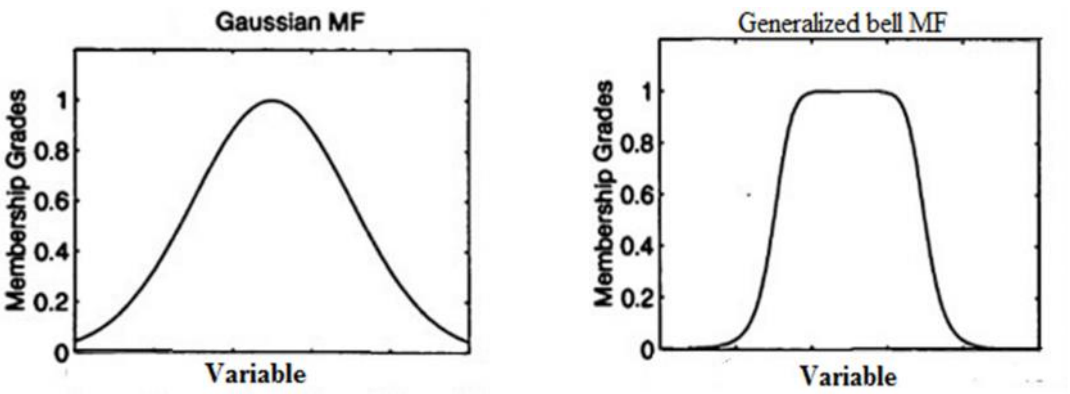
Gaussian MFs is determined completely by c and σ; c represents the MFs Centre and σ determines the MFs width and is presented in Fig. 5.0.

(b) Generalized bell membership functions (MFs)

A generalized bell membership functions (MFs) is specified by three parameters (a, b, c) as:

$$g\,(x; a, b, c) = \frac{1}{1+|(x-c)/a|^{2b}} \quad \text{…. (13)}$$

Where the parameter b is usually positive (If b is negative, the shape of this MF becomes an upside-down bell). This membership functions (MFs) is a direct generalization of the Cauchy distribution used in probability theory, so it is also referred to as the Cauchy MF and is shown in Fig. 5.0.



0

FIGURE 5.0 Various membership functions (MFs) of fuzzy sets

Multiple linear regressions (MLR)

Regression model is another highly recognized method for hydrological prediction. A regression model that involves more than one independent variable is called Multiple Linear Regression Model (MLR). It is a linear relationship between inputs and output. Regression analysis studies the correlation between dependent and independent variables. The major advantage is that it is very straight forward to see how dependent variable change when independent variable differ. Regression analysis finds formula that relates dependent variable and independent variables by the fitting at a linear and non-linear curve to observe data.

Multiple linear regressions are the extended forms of simple linear regressions applied to the case of multiple explanatory variables. The purpose of MLR is to explain as much as possible of the variation observed in the response variable, leaving as little variation as possible to unexplained "noise" (Helsel and Hirsch 2002).

The general form of MLR is given as follows:

$$Y = b_0+b_1X_1+b_2X_2+ \ldots\ldots b_pX_p \quad \text{…. (14)}$$

Where Y is a dependent variable, X_1, X_2, …. X_p are independent variables, b_0 is intercept and b_1, b_2, ….. b_p are regression coefficients.

Gamma Test

GT is one of the non-linear modelling tools whereby an appropriate combination from input parameters can be investigated for modelling the output data as well as establishing a smooth model. GT estimates the minimum mean square errors which is obtainable in continuous non-linear models with unseen data. Suppose there is a set of data as the following:

$$\left((x_1 \ldots x_n)y\right) = (X, y) \quad \text{….. (15)}$$

Where $X = (x_1 \ldots x_n)$ is the input vector in the output vector's areas of y and $C \in R^n$. If the relationship is established between the set members:

$$y = f(x_1 \ldots x_n) + r \quad \text{…. (16)}$$

in which r is a random variable. GT is an estimate for the output variance of a non-smooth model. According to K [i, k], Gamma Test includes a list of k ($1 \le k \le p$) the k^{th} neighbor for each vector X ($1 \le i \le M$). Delta function calculates the mean squared distance of the k^{th} neighbor.

$$\delta_M(k) = \frac{1}{M}\sum_{i=1}^{M}\left|[X_{n[i.k]} - X_i]\right|^2 \quad \text{…. (17)}$$

In which | | indicates Euclidean distance, corresponding gamma function is as:

$$\gamma_M(k) = \frac{1}{2M}\sum_{i=1}^{M}[Y_{n[i.k]} - Y_i]^2 \quad \text{…. (18)}$$

Where $Y_{N[i,k]}$ is the value of y corresponding to the k^{th} neighbor of X_i in the equation 15. In order to calculate gamma the linear regression is fitted from p spot to values of $\delta_M(k)$ and $\gamma_M(k)$.

$$\gamma = A\delta + gamma \quad \text{…. (19)}$$

The intercept of this line $\delta = 0$ indicates the gamma value and $\gamma_M(k)$ is equal to the errors variance. Provided that n is the number of the input variables, the combination $2^n - 1$ of would be among them. Reviewing all these combinations takes a lot of time. GT can identify the most effective variable in modeling and the best combination of the input variables. In addition, M test can also identify the length of training period of the prediction model to establish a smooth model.

PERFORMANCE EVALUATION OF MODELS

Root mean square error (RMSE)

The root mean square error is used to measure the prediction accuracy of a model. It compares difference between predicted and observed values and gets the information on short term performance. It is a positive value ranging from 0 to ∞. The RMSE is zero for perfect fit and increased values indicate higher deviation between predicted and observed values. The root mean square error (RMSE) is determined by following relationship:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Q_o - Q_p)^2}{n}} \quad \text{…. (20)}$$

where, Q_o is i^{th} observed values of daily runoff, Q_p is predicted values of daily runoff and n is the number of observations.

Correlation coefficient (r)

The Correlation Coefficient (r) is an indicator of degree of closeness between observed and predicted values and provides the level of variance explained between observed and predicted. If observed and predicted values are completely independent, the r will be zero (Mutreja, 1992). The correlation coefficient is determined using the following equation:

$$r = \left[\frac{\sum_{i=1}^{n}\{(Q_o - \bar{Q}_o)(Q_p - \bar{Q}_p)\}}{\sqrt{\sum_{i=1}^{n}(Q_o - \bar{Q}_o)^2}\sqrt{\sum_{i=1}^{n}(Q_p - \bar{Q}_p)^2}}\right] \quad \text{…. (21)}$$

Where $\bar{Q}_o$ is average of the observed daily runoff series and $\bar{Q}_p$ is average of predicted daily runoff series.

Coefficient of efficiency (CE)

The Coefficient of efficiency is developed by Nash and Sutcliffe in 1970. It is providing the proportions of variance of the observation for model and used very commonly in hydrology. Nash-Sutcliffe coefficient of efficiencies range between $-\infty$ to 1. The Coefficient of efficiency is one for perfect match between observed and predicted values. Similarly CE value equal to zero indicates that the model predictions are equal to mean of observed data series. The Coefficient of efficiency is determined by using the following equation:

$$\boldsymbol{CE} = 1 - \frac{\sum_{i=1}^{n}(Q_o - Q_p)^2}{\sum_{i=1}^{n}(Q_o - \bar{Q}_o)^2} \quad \text{…. (22)}$$

RESULTS & DISCUSSION

Model input selection

In order to simulation runoff by ANNs, variables Rt, Rt-1, Rt-2, Rt-3, Rt-4, Qt-1, Qt-2, Qt-3 and Qt-4 were considered as input variables. The finding best input combination is the utmost important step of any modeling. As indicated the complexities of the model including the higher number of inputs, more data for training model, a model with greater parameters, may have less prediction error; however, it not necessarily ensure fewer errors at the test phase. In this condition, there is an optimal condition in which prediction errors are minimized at the test phase. Also GT was used for identifying the best input combination of input variables. Different combinations of input variables were explored to assess their influence on the runoff simulation (Table 3.0). Gamma test predicts the minimum achievable modeling error before the modeling. Suppose ‘n’ is the variables influencing on occurrence of a phenomenon; 2^n-1 meaningful combination would be established from the input variables. As indicated in Table 3.0, out of 9 parameters, Rt, Rt-1, Qt-1 had the highest influence on runoff discharge. Moreover, eliminating the parameters Rt-3, Rt-4, Qt-4 decreased the gamma value. Eliminating other remained variables had an identical influence on increasing the gamma value.

TABLE 3.0 Identifying the most effective variable based on GT

Different combinations	Mask	Gamma	SE	V_{ratio}
All	111111111	0.0322	0.0143	0.1288
All - R_t	011111111	0.0501	0.0129	0.2006
All - R_{t-1}	101111111	0.1142	0.0151	0.4570
All -R_{t-2}	110111111	0.0393	0.0156	0.1574
All - R_{t-3}	111011111	0.0257	0.0101	0.1030
All - R_{t-4}	111101111	0.0317	0.0115	0.1350
All - Q_{t-1}	11110111	0.0381	0.0132	0.1524
All - Q_{t-2}	11111011	0.0347	0.0193	0.1391
All - Q_{t-3}	11111101	0.0345	0.0195	0.1502
All - Q_{t-4}	111111110	0.0252	0.0183	0.1008

To determine the best input combination in modeling, various combinations of input parameters were assessed using GT so as to identify the most appropriate combination among the remained variables to predict the runoff discharge. Some of these combinations along with Gamma values are shown in Table 4.0. The results showed, the best input combination of the variable is when using Rt, Rt-1, Rt-2, Qt-1. The Small value of gamma indicates that the data with provided combination might possibly provide better results in modeling.

ANFIS and MLR model results

The ANFIS and MLR models with best input were compared based on their performance in training sets and testing period. The results were summarized in Tables 5.0.

It was apparent that all of the performances of these models are similar during training as well as testing. It also showed that the model, which consisted of two lag days rainfall and one lag days runoff data in input, had the smallest value of the RMSE as well as higher value of r and CE in the training as well as testing period, so, it was selected as the best-fit model for predicting the river flow in this study. In order to get an effective evaluation of the ANFIS and MLR models performance, the best model structures, has been used to compare the ANFIS and MLR models. From the best fit model, it was found that the difference between the values of the statistical indices of the training and testing set does not vary substantially. It was observed that all models generally gave low values of the RMSE and as well as high r and CE, the performances of the ANFIS and MLR models performance in the runoff forecasting were satisfactory. Analyzing the results during training, it can be observed that the ANFIS model trained using Gaussian MFs and ANFIS-2 model trained using Generalized bell MFs. Analyzing the results from the table 5.0. It can say that ANFIS-1 model having the best statistical result during training and testing period in compared to ANFIS-2 model. The results of the MLR model were obtained the best RMSE, r and CE statistics of 36.42, 0.76 and 0.78 85 respectively during the training period, Similarly obtained the best RMSE, r and CE statistics of 26.66, 0.80 and 0.85 respectively during the testing period.

The qualitative performance of developed model was judged by observed and predicted daily runoff hydrograph and scatter plots. Fig. 6.0-7.0 showed the hydrograph and scatter plots of both the observed data and the predicted obtained by using the ANFIS-1, ANFIS-2 and MLR model during testing period. It was visibly seen from the hydrographs and scatter plots that the ANFIS-1model estimates were closer to the corresponding observed flow values than those of the other models. As seen from the fit line equations (assume that the equation is y = ax + b) in the scatter plots that a and b coefficients for the ANFIS-1 model are, respectively, closer to the 1 and 0 with a higher R value of 0.95 than ANFIS-2 and MLR models. The models of ANFIS-1 and MLR showed good prediction accuracy for low values of runoff but were unable to maintain their accuracy for high values of runoff. However, a significant improvement is observed for the ANFIS-2 in the peak runoff prediction compared to ANFIS-1 and MLR.

Overall, the ANFIS-1, ANFIS-2 and MLR models can give good prediction performance and could be successfully applied to establish the predicting models that could provide accurate and reliable daily runoff prediction. The results suggest that the ANFIS model was superior to the MLR in the runoff prediction.

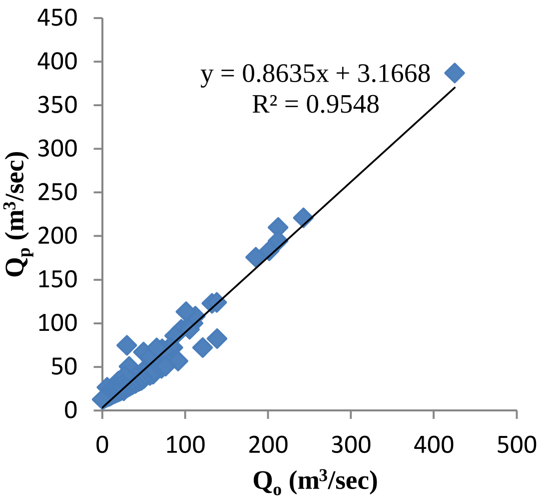
The results were also tested by using t-test for verifying the robustness (the significance of differences between the model observed and predicted values) of the optimal ANFIS-1, ANFIS-2 and MLR models. Test was set at a 95% significant level. The statistics of the tests are provided in Table 6.0. The ANFIS-1 model yields smaller testing values with a higher significance level than the ANFIS-2 and MLR models. According to the test result, the ANFIS-1 model seems to be more robust (the similarity between the observed runoff values and ANFIS-1 estimates are high) in rainfall-runoff modeling than the other models.

TABLE 5.0 Statistical performance evaluation measures from various ANFIS and MLR model

Model	Networks	Training			Testing		
		RMSE	r	CE	RMSE	r	CE
ANFIS-1	Gauss(5)	18.12	0.97	0.96	10.04	0.98	0.97
ANFIS-2	Bell(4)	19.43	0.97	0.95	11.57	0.97	0.96
MLR		36.42	0.76	0.78	26.23	0.80	0.85

TABLE 6.0 t-test results of ANFIS and MLR models

Model	t-statistic	Resultant significance level
ANFIS-1	0.03941	0.9764
ANFIS-2	0.04512	0.9671
MLR	1.9756	0.2145



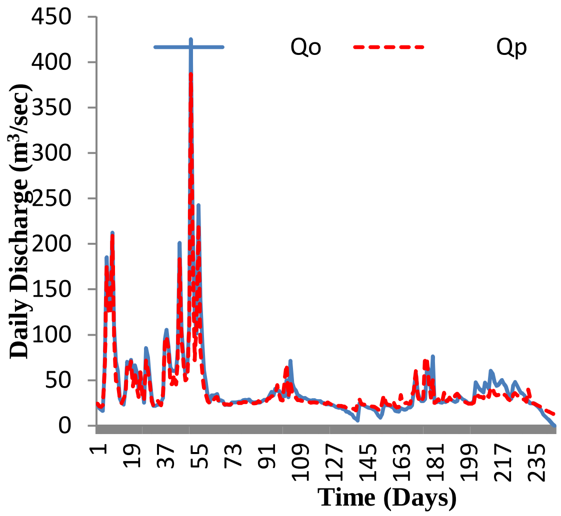


FIGURE 6.0 Comparison of observed (Qo) and predicted (Qp) daily runoff and their corresponding scatter plot during testing period for ANFIS-1 model

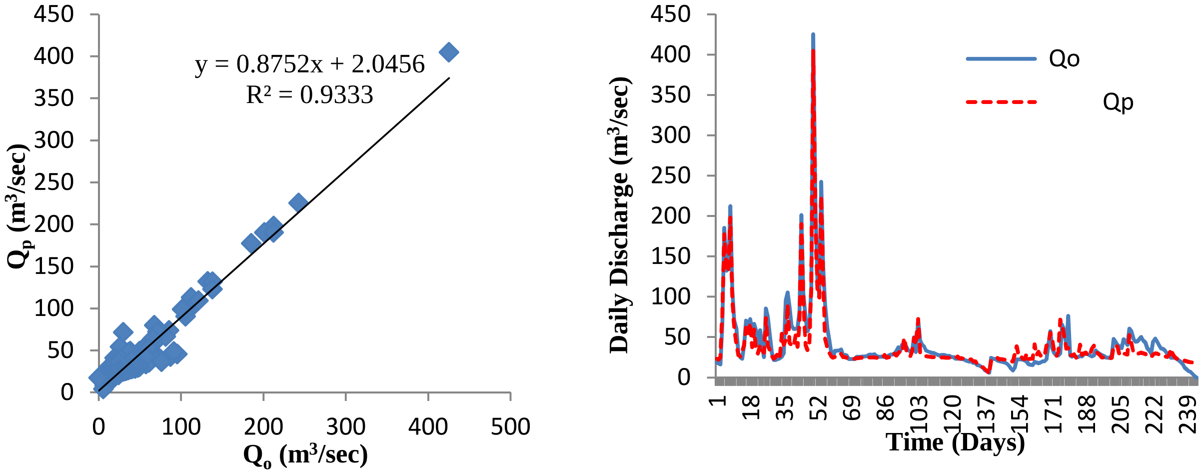


FIGURE 7.0 Comparison of observed (Qo) and predicted (Qp) daily runoff and their corresponding scatter plot during testing period for ANFIS-2 model

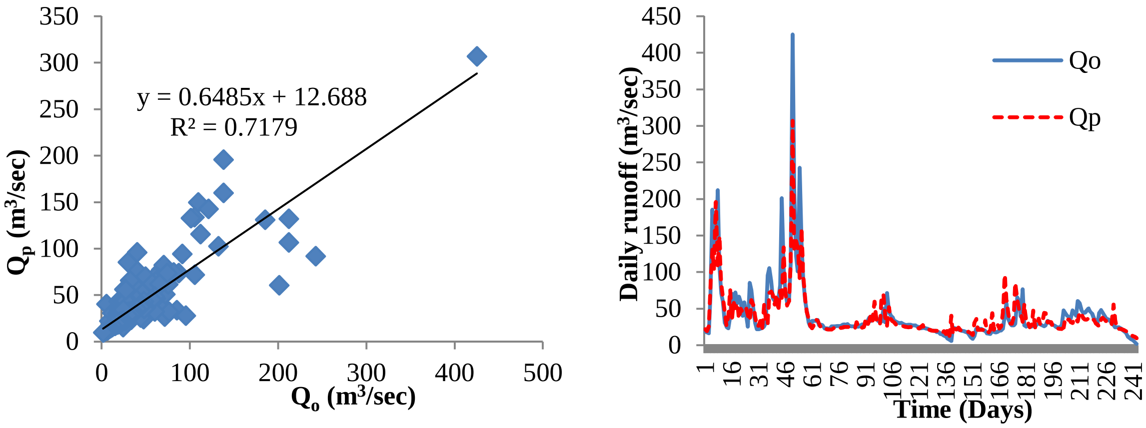


FIGURE 8.0 Comparison of observed (Q_o) and predicted (Q_p) daily runoff and their corresponding scatter plot during testing period for MLR model

Some researchers have reported that the ANN rainfall–runoff models trained using popular BPA do not perform well in predicting low magnitude de flows (Jain and Srinivasulu, 2006). In order to compare the performances of ANFIS models viz. ANFIS-1 and ANFIS-2 and MLR model, for this, selected error statistics (RMSE, r and CE) were calculated from different models for the data corresponding to low and high magnitudes of flow. The results of this analysis are presented in Table 7.0. It can be noted from Table 6.0 that during training and testing period ANFIS-1 model with values of statistics RMSE, r and CE of 09.14, 0.94 and 0.96 for low magnitude flows. It was concluded that ANFIS-1 model trained using Gaussian MFs having best performing results during low magnitude flow. Similarly during high magnitude flow the ANFIS-2 model was found best results of statistics of RMSE, r and CE of 20.14, 0.99 and .99 respectively during testing period.

TABLE 7.0 Statistical results for low and high magnitude flows

Model	Training			Testing		
	Low magnitude flows					
	RMSE	r	CE	RMSE	r	CE
ANFIS-1	52.01	0.82	0.87	09.14	0.94	0.96
ANFIS-2	51.58	0.81	0.86	10.51	0.93	0.94
MLR	32.12	0.63	0.59	20.15	0.72	0.81
High magnitude flows						
ANFIS-1	61.27	0.98	0.98	31.29	0.97	0.98
ANFIS-2	54.12	0.98	0.98	20.14	0.99	0.99
MLR	192.21	0.66	0.78	91.41	0.68	0.79

After the analyzing the results both during training and testing, it was found that ANFIS-1 model having best performance in low magnitude flow, but during high magnitude flow the ANFIS-2 model was found to be the best performing results. Improvements in the r and CE statistics also can be noted from Table 7.0 for all magnitude flows during both training and testing data sets by the ANFIS models having the best result in compared to MLR models. The visual result of best model during low flow and high flow shown in Fig. 9.0 and 10.

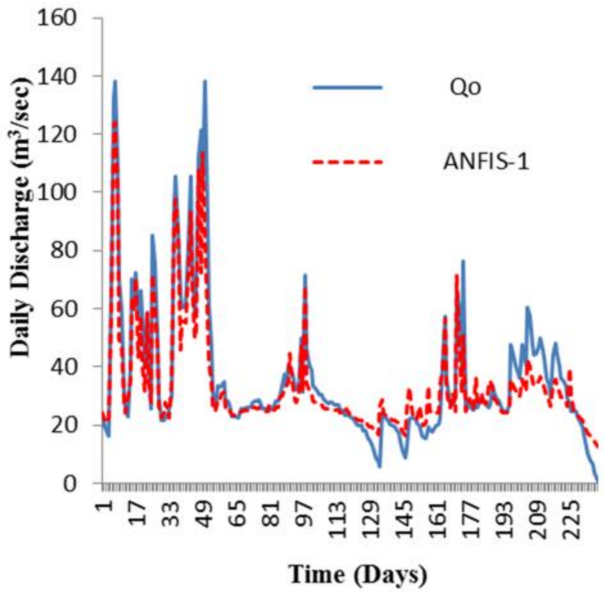


FIGURE 9.0 Observed and predicted runoff during low flow magnitutite

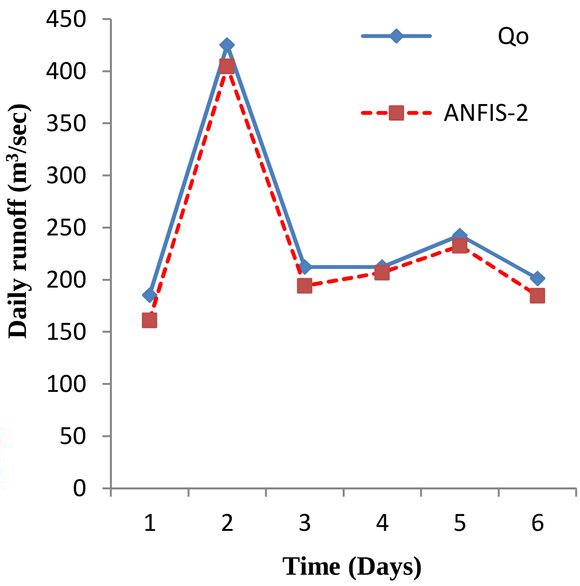


FIGURE 10 Observed and predicted runoff during high flow magnitutite

CONCLUSION

In this study, ANFIS and MLR models were developed for prediction the short term of runoff based on antecedent values of runoff data. For achieving this objective, the Ghatora station located in the Arpa River has been selected as case study. The results of ANFIS and MLR models and observed values were compared and evaluated based on their training and testing performance. The results demonstrated that ANFIS and MLR can be applied successfully to establish accurate and reliable river flow forecasting. According to Gamma test the results; the model which consists of on antecedent values of rainfall and runoff has been selected as the best fit prediction input variable. Comparing the results of ANFIS and MLR models, it was seen that the values of R and CE of ANFIS models were higher than those of MLR models. Also, the RMSE values of ANFIS models were lower than those of MLR models. Therefore, ANFIS model could improve the accuracy over the MLR models. The results also demonstrated ANFIS showed good prediction accuracy for low values of flow and high values of flow. However, a significant improvement is observed for the ANFIS in the peak flow prediction compared to MLR. Overall, the analysis presented in this study provides that the ANFIS method was superior to the MLR in the runoff prediction and Gaussian membership function (MFs) was superior performance than to generalized bell membership function (MFs) for training of ANFIS model in the study area.
Although the results presented here are promising and these data driven models can be successfully applied to establish runoff with complicated topography forecasting models, these models underestimate significantly flow in the flood conditions. In future work, further research is necessary to improve the prediction accuracy, especially for the high.

REFERENCES

Asati, S.R. and Rathore, S.S. (2012) Comparative study of stream flow prediction models. Int. Jou. Life Sci. Biotechnology Pharm Res 1(2): 139–151

Bisht, D.C.S., Raju, M.M. and Joshi, M.C. (2010) ANN based River stage-discharge modeling for Godavari River, India. Comput Model New Technol 14(3): 48–62.

Bisht, D.C.S. and Jangid, A. (2011) Discharge modelling using adaptive neuro-fuzzy inference system. International Journal of Advanced Science and Technology, Vol. 31.

Dawson, C.W. and Wilby, R.L. (1999) A comparison of artificial neural networks used for River flow forecasting. Hydrol Earth Syst Sci 3(4): 529–540.

Folorunsho, J. Garba, S. Obiniyi, A.A. and Ajibade, A.O. (2014) A comparison of ANFIS and ANN–based models in River discharge forecasting, New Ground Research Journal of the Physical Sciences, Vol1 (1): 1-16.

Ghose, D.K. Panda, S.S. & Swain, P.C. (2013) Prediction and optimization of runoff via ANFIS and GA. Alexandria Engineering Journal, 52: 209–220

He, Z., Wen, X., Liu, H. and Dua, J. (2014) A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting River flow in the semi-arid mountain region. Journal of Hydrology, 509: 379–386.

Helsel, D.R. and Hirsch, R.M. (2002) Statistical Methods in Water Resources. Techniques of Water- Resources Investigations of the U. S Geological Survey Book 4, Hydrologic Analysis and Interpretation, Chapter A3.

Jain, A. and Srinivasulu, S. (2006) Integrated approach to modelling decomposed flow hydrograph using artificial neural network and conceptual techniques, J. Hydrol., 317(3-4): 291-306.

Jang, J.S.R. (1993) An adaptive-network-based fuzzy inference system. IEEE Transactions on Systems, Man and Cybernetics 23: 665–685.

Kermani, M. Z. and Teshnehlab, M. (2007) Using adaptive neuro-fuzzy inference system for hydrological time series prediction. Hydrological Sciences Journal, 47(6): 1568-4946.

Loukas (2001) Adaptive neuro-fuzzy inference system: an instant and architecture-free predictor for improved QSAR studies. J. Med. Chem, 44(17): 2772–2783.

Mutreja, K.N. (1992) Applied Hydrology. Tata McGraw-Hill, Pub. Co. Ltd., New Delhi.

Nash, J. E., Sutcliffe, J.V. (1970) River flow forecasting through conceptual models. Part 1- A discussion of principles. Journal of Hydrology, 10: 282-290.

Nayak P.C. Sudheer, K.P. Rangan, D.M. and Ramasastri, K. (2003) Short-term flood forecasting with a neuro-fuzzy model. Water Resour. Res., doi: 10.1029/2004 WR0 03562.

Nayak, P.C., Sudheer, K.P., Rangan, D.M. and Ramasastri, K.S. (2004) A neurofuzzy computing technique for modelling hydrological time series. Journal of Hydrology, 291: 52-66.

Panchal, R. Suryanarayana, and Parekh, T.M.V. (2014) Adaptive Neuro-Fuzzy Inference System for rainfall-runoff modeling. Int. Journal of Engineering Research and Applications www.ijera.com Vol. 4, ISSN: 2248-9622.

Rezaeianzadeh, M. Tabari, H. Yazdi, A.A. Isik, S. and Kalin, K. (2013) Flood flow forecasting using ANN, ANFIS and regression models. Neural Comput & Applic DOI 10.1007/00521-013-14436.

Shamseldin, A.Y. O'connor, K.M. and Nasr, A.E. (2010) A comparative study of three neural network forecast combination methods for simulated River flows of different rainfall-runoff models. Hydrological Sciences Journal. 52(5): 896-916.

Shrivastav, M.B. Gandhi, H.M. Ramanuj, P.K. Chudasama, M.K. and Joshi, J.A. (2014) Prediction of Runoff using Artificial Neural Networks (A Case study of Khodiyar Catchment Area). International Journal for Scientific Research & Development Vol. 2 ISSN-2321-0613

Talei, A., Chua, L.H.C. and Wong, T.S.W. (2010) Evaluation of rainfall and discharge inputs used by Adaptive Network-based Fuzzy Inference Systems (ANFIS) in rainfall–runoff modeling. Journal of Hydrology, 391: 248–262.

Tokar, A.S. Johnson, P.A. (1999) Rainfall runoff modeling using artificial neural network. Journal of Hydrologic Engineering, ASCE 4 (3): 232–239.

Xiong, L.H. Shamseldin, A. Connor, O. and K.M. (2013) A non-linear combination of forecasts of rainfall–runoff models by the first-order TS fussy system for forecast of rainfall–runoff model. Journal of Hydrology, 245: 196–217.