Journal homepage www.jzs.univsul.edu.iq
Journal of Zankoy Sulaimani
Part-A- (Pure and Applied Sciences)

Prediction of Darbandikhan Reservoir Inflow Using ANFIS Models

Dr. Hekmat M. Ibrahim
Faculty of Engineering-Sulaimani University, Bakrajo Street, Sulaimaniyah-Iraq
Email: hekmat.ibrahim@univsul.edu.iq

Article info

Original: 1 October 2016
Revised: 23 December 2016
Accepted: 9 January 2017
Published online: 20 June 2017

Key Words:
Darbandikhan
Reservoir
Inflow
Prediction
ANFIS

Abstract

Predicting reservoir inflow is important to reservoir management and scheduling; therefore, accurate prediction of reservoir inflow has been a vital task for researchers and water resources managers during last decades. Recently, Adaptive Neuro-Fuzzy Inference System (ANFIS) has been extensively used to find the relationship between inputs and outputs without considering the physical process of the phenomena. Therefore, the current study investigates and evaluates the capability and applicability of the ANFIS technique in modeling the reservoir inflow for predicting future values of monthly inflow to Darbandikhan reservoir in Kurdistan Region, Iraq. Five rain gage stations data from 1992-2009 were used to develop ANFIS models to predict monthly reservoir inflow from various combinations of antecedent values of inflow and rainfall within the basin in Iraq and Iran which are considered as input variables. Two different membership functions (MFs), triangular and generalized bell, with different numbers (2-5) for each input variable were investigated in the ANFIS models. The best fit models were selected using three performance evaluation criteria, namely; coefficient of determination (R^2), coefficient of efficiency (CE) and normalized root mean squared error ($NRMSE$).

In comparison of ANFIS model uses the triangular MF with that model uses the generalized bell MF, approximately, similar predict accuracies were obtained for both MFs except the latter needs fewer numbers of MFs. The results indicated that using a combination of inflow and rainfall as input variables is effective in improving predict accuracy and developing parsimonious models with fewer and readily available input variables. Moreover, among different architectures of ANFIS, structures, including three input variables, 1 and 12 time lags of inflow (I_{t-1}, I_{t-12}) and rainfall of Marivan station (Rm_t) showed better performance for this application. For best models, the performance evaluation criteria, R^2, CE and $NRMSE$, for checking data set were obtained as 0.96, 0.95 and 0.2 respectively for model with four triangular MFs; and as 0.96, 0.96 and 0.2 respectively for model with three generalized bell MFs. Reservoir inflow modeling in this way will be more reliable than doing it using a time series model as a more effective parameter could be incorporated. By considering the results, ANFIS method is an effective tool that can be successfully applied for reservoir inflow modeling and have satisfactory performances in Darbandikhan reservoir monthly inflow prediction.

Introduction

Reservoirs are the most important and effective water storage facilities for distributing water among different demands such as drinking, irrigation and hydroelectric energy and also smooth out extreme inflows to mitigate droughts or floods. To make the best use of the available water resource and help to reduce the possible problems downstream, the optimal reservoir operation in a system is very important and depends on three main parameters, inflow, storage and downstream demand. Reservoir inflow involves the intrinsic uncertainty and imprecision characteristics in the decision-making process and operating policy of the reservoir. Hence, increasing the accuracy of reservoir inflow predicting has prompted great interest in water resources engineering.

The study of the hydrologic time series predicting in the past several decades has produced great exhilaration and care, and a huge number of models and approaches have been proposed to improve the accuracy of predicting. The models can be divided into statistical methods, physical methods, and intelligent approaches. However, there was no one predicting method that was proper, generally, for any reservoirs inflow because of the changing in the hydrological characteristics of river basins and regions with variation of time and space, and each category of method has its various advantages and deficiencies. Various unstable factors and always present such as characteristics time varying, nonstationary, and significant outliers affect the hydrological data. The correlation between the past and future values is changed by the characteristics of time series data. Moreover, different time series regions have many noise levels, which further increase the difficulty of predicting models. Therefore, it is difficult for a single time series predicting model to capture the dynamic changing processes and features, which may run into local under fitting or over fitting problems. In order to increase models performance, researchers have been continuously developing new technologies and methods for the hydrological prediction. In recent years, many hybrid approaches take advantage of more than one predicting method to carry out the research work and engineering practice in different fields. Application results indicate that the hybrid methods have been higher predicting precision than a single predicting method [1] and [2].

The Adaptive Neuro-Fuzzy Inference Systems (ANFIS) is one of the hybrid approaches that combine fuzzy logic and neural network technology as it provides an accurate and powerful alternative in modelling numerous processes. More recently, literatures have found the application of ANFIS in many fields, such as, regional electricity loads, ophthalmology, reservoir operation, wind speed, evaporation, river flow prediction, etc. [3], [4], [5] and [6]. Many successful applications demonstrate that, with the advantages of good generality and predict accuracy; ANFIS is an efficient and promising approach in hydrological prediction, particularly in reservoir inflow prediction. Water inflow to the reservoir is affected by many factors such as rainfall, temperature, etc., and a reliable predicts should normally consider the effects of such factors [7], [8] and [9].

Hence, the main purpose of this study is to develop an accurate model based on the ANFIS technique for reservoir monthly inflow prediction. High performances have been obtained when the technique used with monthly inflow prediction for Darbandikhan reservoir in Kurdistan Region, Iraq. The performance of the models investigated for training and checking data sets are compared to the observations, and the best fit model is identified according to the performance criteria, including the coefficient of determination (R^2), coefficient of efficiency (CE), and normalized root mean square error ($NRMSE$).

Study Area and Used Data

A. Study Area

The data sources critical to the development of any comprehensive predicting model. Because the development of the reservoir inflow predicting models in the present study is dependent on the availability of water inflow to reservoir and rainfall time series data, therefore, to illustrate the applicability and capability of the models developed are applied to area, which can access to the required data. The areas that the required data have been obtained for, and the developed models applied to it as a case study are Darbandikhan basin and reservoir.

Darbandikhan reservoir is a large freshwater reservoir impounded by the Darbandikhan dam which is located on the Diyala-Sirwan river in Kurdistan Region, Iraq approximately 65 km south-east of Sulaimani and 230 km north-east of Baghdad. The reservoir is fed by three rivers (Tanjero, Zalm and Sirwan) and their tributaries in Darbandikhan basin, which is distributed between Iraq and Iran as shown in Figure-1 [10]. Many people benefit from Darbandikhan reservoir and its waters for drinking, irrigation, fisheries, recreation and electricity generating by the Darbandikhan dam [11].

Darbandikhan reservoir had an area of 113 km² at normal operating level (El. 485.0 m ASL) and a total design capacity of 3×10^9 m³, of which 2.5×10^9 m³ is live storage and 0.5×10^9 m³ being dead storage. The current storage volumes will be less than this due to 55 years of sedimentation. The catchment area of the reservoir is about 17,850 km² with a mean annual catchment precipitation of 840 mm and maximum recorded inflow of 5816 m³/s [12].

B. Used Data

According to the availability of data, the monthly data of inflow into the reservoir of Darbandikhan dam and recorded rainfall from five rain gage stations within the basin from January, 1992 to December, 2009 (18 years, 216 months) were used as inputs to the developed ANFIS models. Two of the rain gages are located in Iraq (Sulaimani and Darbandikhan), and the other three stations are located in Iran (Marivan, Sanandaj and Kermanshah) as shown in Figure-1. The monthly time series data of inflow to reservoir and rainfall of Darbandikan gage station were obtained from the Directorate of Darbandikhan dam while the rainfall of Sulaimani gage station was obtained from Sulaimani meteorological center. On the other hand, the monthly time series data of rainfall of Marivan, Sanandaj and Kermanshah gage stations in Iran were downloaded from Islamic Republic of Iran Meteorological Organization (IRIMO) website [13]. Figure-2 represents the time series data for inflow to Darbandikhan reservoir and rainfall at five gaging stations that used in this study.

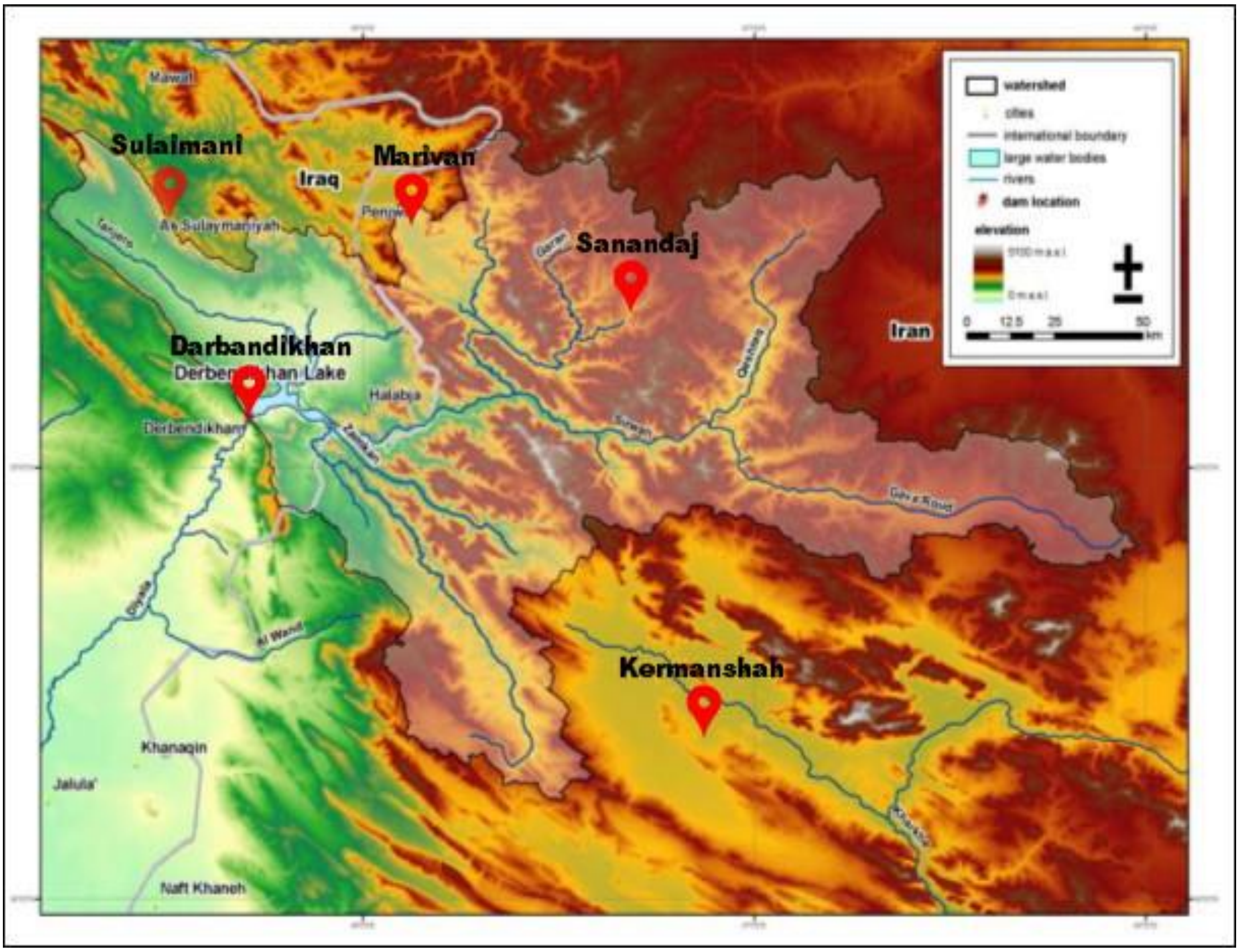

Figure-1: The basin of Darbandikhan showing the locations of reservoir and rain gage stations.

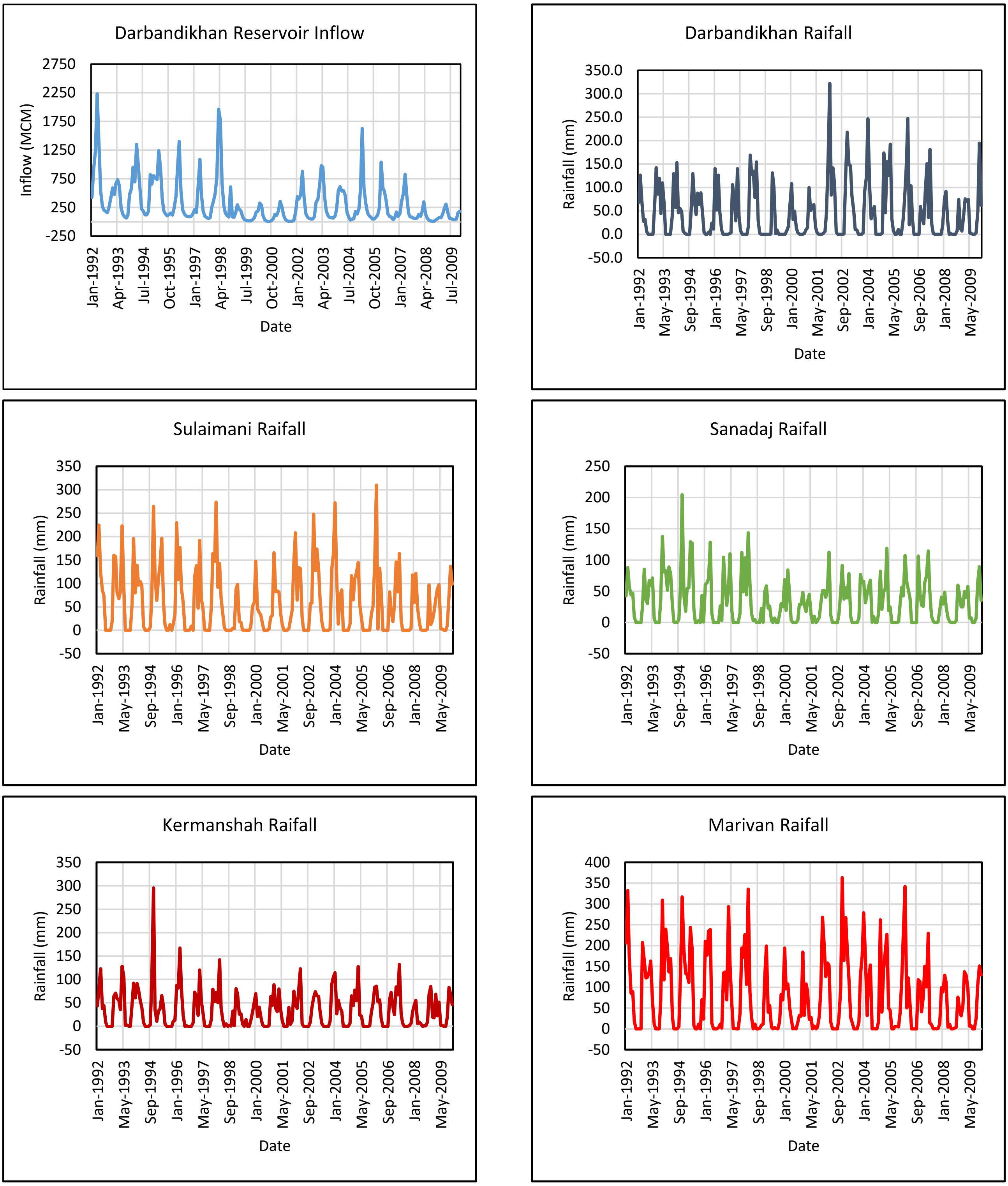

Figure-2: Observed values of Darbandikhan reservoir inflow and rainfall at different stations in Iraq and Iran.

Models Evaluation Criteria

The most important criterion for evaluating predicting models or choosing between competing models is accuracy. Generally speaking, the more accurate predicting model is the closer the predicted values to the observed values of the time series. The performances of the developed models in the current study are evaluated according to statistical evaluating criteria such as the coefficient of determination (R^2), coefficient of efficiency (CE), root mean squared error ($RMSE$) and normalized root mean squared error ($NRMSE$). The definition of these performance criteria are given as follows:

$$R^2 = \left(\left(\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \tilde{y})\right) \Big/ \left(\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2(\hat{y}_i - \tilde{y})^2}\right)\right)^2 \quad \dots\dots\dots\dots \quad (1)$$

$$CE = \left(\sum_{i=1}^{N}(y_i - \bar{y})^2 - \sum_{i=1}^{N}(y_i - \hat{y}_i)^2\right) \Big/ \left(\sum_{i=1}^{N}(y_i - \bar{y})^2\right) \quad \dots\dots\dots\dots \quad (2)$$

$$RMSE = \left(\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2\right)^{0.5} \quad \dots\dots\dots\dots \quad (3)$$

$$NRMSE = \left(\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2\right)^{0.5} \Big/ \left(\frac{1}{N}\sum_{i=1}^{N}(y_i)\right) \quad \dots\dots\dots\dots \quad (4)$$

In which N is the number of data points, y_i, $\bar{y}$ are the observed data with its mean respectively and $\hat{y}$, $\tilde{y}$ are the corresponding predicted data with its mean respectively. The coefficient of determination (R^2) is commonly used statistical criteria and provides information on the strength of linear relationship between the observed and predicted values. The coefficient of efficiency (CE) is statistical criteria employed to evaluate model performance. Values of R^2 and CE close to 1.0 indicate good model performance. The root mean squared error ($RMSE$) statistical criteria that indicate a quantitative measure of the model error in units of the variable. Model performance is indicated well when RMSE value is close to zero. The normalized root mean squared error ($NRMSE$) is statistical criteria that indicate a model's ability to predict a value away from the mean.

A more realistic way of assessing a model's accuracy is to use a checking data set; that is, some of the data at the end of the series are omitted before the models are estimated. Then the models are compared on the basis of how well they predict the data, which have been withheld rather than how well they predict the same data which has been used for modeling [14]. Therefore, in the present study for selecting and comparison of predicting models, the time series data were split into two data sets. The first set of the data is used to build the predict model (training data set). Then the model is used to predict ahead the remaining data points (checking data set). To obtain information about the model's ahead predicting performance; the resulting out of the model predicts is compared to the observed series values. Some of the above measures are used for checking data set and models that yield best values for these statistics on checking data set, would be chosen as a good model. For monthly Darbandikhan reservoir inflow data, the last 18 months of the data were used as checking data set, then the models were fitted on the remaining data and used to predict 18 months ahead. According to this partitioning of data, the number of training and checking data sets will be 198 and 18 months respectively. At the end, the best two models were examined by using 6, 12, 18 and 24 months as checking data set and used to predict 6, 12, 18 and 24 months ahead.

ANFIS Methodology

A neuro-fuzzy system is defined as a combination of neural networks and fuzzy inference system. Jang in 1993 introduced an adaptive neuro-fuzzy inference system (ANFIS) where the membership functions (MFs) parameters are fitted to a data set through a hybrid learning algorithm [15]. ANFIS eliminates the problem of defining the membership function (MF) parameters and design of if-then rules in fuzzy system design by using the learning capability of ANN for automatic fuzzy rule generation and parameter optimization [5]. For this reason, in this study, with aids of MATLAB software, the ANFIS methodology is proposed to self-organize model structure and to adapt parameters of the fuzzy system for predicting monthly inflow to Darbandikhan reservoir. It has the advantage of allowing the extraction of fuzzy rules from numerical data.

There are two types of fuzzy inference systems, Sugeno-Takagi and Mamdani. The Sugeno-Takagi type is available in Matlab ANFIS toolbox only. Therefore, in this study, it is used for predicting Darbandikhan reservoir inflow. The ANFIS incorporates five-layer network to implement a Takagi-Sugeno-type fuzzy system as shown in Figure-3, which is considered for simplicity as a fuzzy system with only two inputs and one output. The output of each layer is the input of the next layer. The model has a relatively complex architecture for a large number of inputs, and it can process a large number of fuzzy rules. The ANFIS uses the least mean square training algorithm in the forward computation to determine the linear consequents of the Takagi-Sugeno rules, while for the optimal tuning of an antecedent MF, backpropagation is used [16]. Figure-3 (a), illustrates the fuzzy reasoning mechanism for the first order Takagi-Sugeno model to derive an output function (f) from a given input (x, y):

$$\left.\begin{array}{lll} Rule\ 1: & If\ x\ is\ A_1\ and\ y\ is\ B_1 & then \quad f_1 = p_1x + q_1y + r_1 \\ Rule\ 2: & If\ x\ is\ A_2\ and\ y\ is\ B_2 & then \quad f_2 = p_2x + q_2y + r_2 \end{array}\right\} \quad \ldots\ldots \quad (5)$$

where f_i is the output, x and y are the crisp inputs to the node i, A_i and B_i are the linguistic labels, (p_i, q_i, r_i) are the consequent parameters, μ_{A_i} and μ_{B_i} are the MFs for A_i and B_i linguistic labels, respectively. Figure-3 (b) shows a typical ANFIS architecture with two inputs, (x, y), and one output, (f). The ANFIS building with all the relationships between the input and output of each five layers are described as follows:

Layer 1 (input nodes): Each node in this layer generates membership grades of an input variable. The node output is defined by O_i^1 is calculated by:

$$O_i^1 = \mu_{A_i}(x) \quad i = 1, 2 \qquad or \qquad O_i^1 = \mu_{B_{i-2}}(y) \quad i = 3, 4 \quad \ldots\ldots \quad (6)$$

In this study, the triangular and generalized bell functions are used as:

$$\left.\begin{array}{ll} O_i^1 = \mu_{A_i}(x) = max\left(\min\left(\dfrac{x - a_i}{b_i - a_i}, \dfrac{c_i - x}{c_i - b_i}\right), 0\right) & triangular\ function \\ O_i^1 = \mu_{A_i}(x) = \dfrac{1}{1 + \left|\dfrac{x - c_i}{a_i}\right|^{2b_i}} & generalized\ bell\ function \end{array}\right\} \quad \ldots \quad (7)$$

where (a_i, b_i, c_i) are the parameter set that changes the shapes of the MFs with the maximum equal to 1 and the minimum equal to 0; and called premise parameters or antecedent parameters. For triangular function, a_i and c_i locate the feet of the triangle and the parameter b_i locates the peak.

Layer 2 (rule nodes): the outputs of this layer, called firing strengths O_i^2, are the products of the corresponding degrees obtained from the layer 1:

$$O_i^2 = w_i = \mu_{A_i}(x).\mu_{B_i}(y) \qquad i = 1, 2 \quad \ldots\ldots \quad (8)$$

Layer 3 (average nodes): main target is to compute the ratio of firing strength of each i^{th} rule to the sum firing strength of all rules. The firing strength in this layer is normalized $\overline{w_i}$ as:

$$O_i^3 = \overline{w}_i = \frac{w_i}{\sum_{i=1}^{2} w_i} \qquad i = 1, 2 \quad \ldots\ldots \quad (9)$$

Layer 4 (adaptive nodes): the contribution of i^{th} rule towards the total output or the model output and the function defined is calculated by:

$$O_i^4 = \overline{w}_i f_i = \overline{w}_i(p_i x + q_i y + r_i) \qquad i = 1, 2 \quad \ldots\ldots \quad (10)$$

Layer 5 (output nodes): this layer is called as the output nodes in which the single node computes the overall output by summing all incoming signals:

$$O_i^5 = f = f(x, y) = \frac{w_1 f_1 + w_2 f_2}{w_1 + w_2} = \overline{w}_1 f_1 + \overline{w}_2 f_2 \quad \ldots\ldots \quad (11)$$

Generally, the methodology of carrying out this study can be summarized as follow:

1. Selecting independent variables based on the data available. In this study, inflows to Darbandikhan reservoir and rainfall of different rain gage stations with their antecedents are used as independent input variables.
2. Building models based on inflow to Darbandikhan reservoir and rainfall of different rain gage stations. The models based on the combination of all variables, inflow and rainfall, with their antecedents.
3. Training and checking the models by ANFIS using a different number of triangular and generalized bell MFs and then selecting the best fit models.

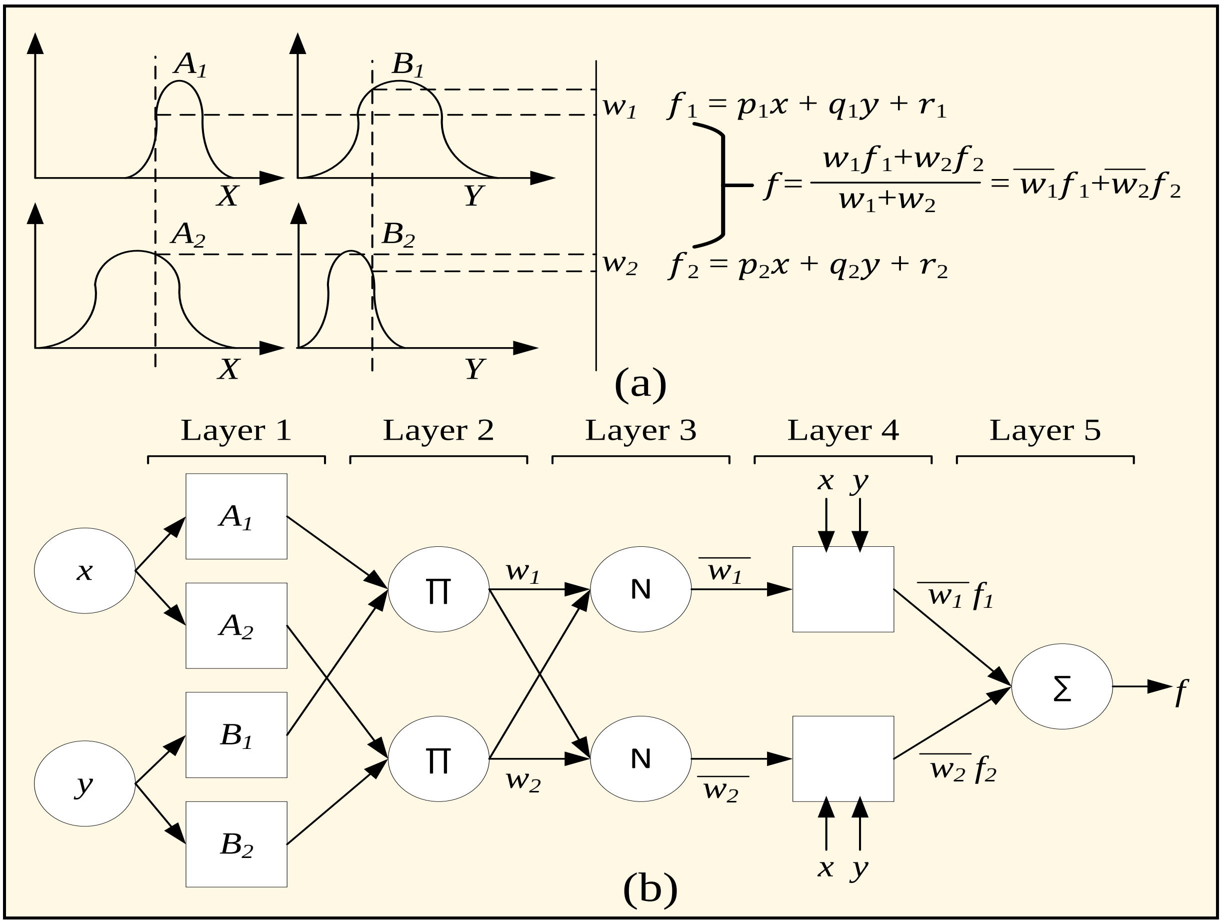

Figure-3: (a) First-order Takagi-Sugeno fuzzy model, (b) Equivalent ANFIS architecture.

Reservoir Inflow Prediction

A. Input Variables Selection

One of the most important steps in the model development process is the determination of significant input variables. Consider modeling a reservoir inflow time series, where it is required to predict the value of inflow I_{t+1} at time $t+i$, where i is the lead time. The inputs to the ANFIS are typically chosen as the values of the time series up to time t and the output will be the predicted value. In addition to previous values of the time series, one can utilize the values or predicts of other time series or external variables as inputs that have a correlated or causal relationship with the series to be predicted. For a reservoir inflow predicting problem such exogenous time series will be the rainfall R in the reservoir basin. The functional form of this type of model is:

$$I_{t+i} = f(I_t, I_{t-1}, \dots, I_{t-j}, R_t, R_{t-1}, \dots, R_{t-k}) \quad \dots\dots\dots\dots\dots\dots\dots\dots \quad (12)$$

where f is the unknown function mapped by the model, i is an index representing lead time, and j and k are the maximum number of time steps in the past considered important in modeling I_{t+i} and R_{t-k} is the exogenous variable (rainfall) considered as an input. In this study, the appropriate values of j and k are determined by using a statistical approach which is based on the heuristic that the potential influencing variables corresponding to different time lags can be identified through statistical analysis of the data time series that uses cross correlation, autocorrelation, and partial autocorrelation between the variables.

The autocorrelation function shows a significant correlation, at 95% confidence level, at 1, 2, 11, and 12 months of inflow lag on the inflow at any time as shown in Figure-4. The cross correlation between the inflow to Darbandikhan reservoir and different stations rainfall series at various lags shows significant correlation at 0, 1, and 2 months of rainfall lag on the inflow at any time as shown in Table-1. Therefore, a total number of 19 variables (4 inflows and 15 rainfall) are identified as input variables as shown in Table-2.

As mentioned earlier, the total 216 monthly data records for each variable were collected in the period 1992–2009. The data set was divided into two subsets, training and checking data set. The training data set includes a total of 198 data records. In order to get more reliable evaluation and comparison, models are tested with checking data set that was not used during the training process. The checking data set consists of a total of 18 data records, which is the last 18 months of the time series.

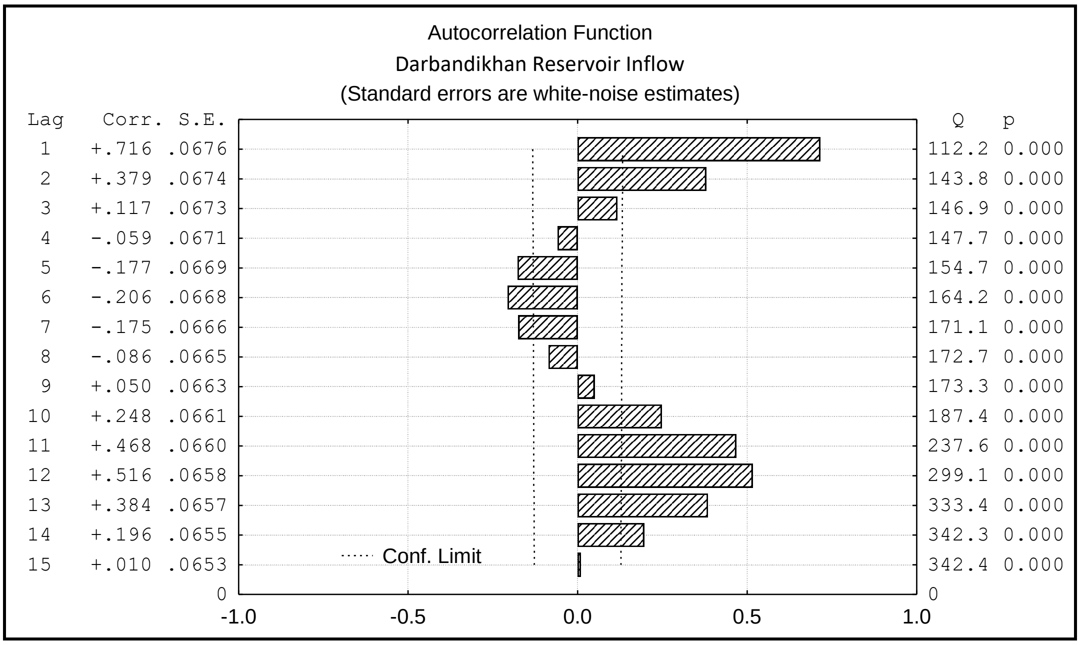

Figure-4: Auto correlation coefficients of Darbandikhan reservoir inflow (I).

Table-1: Cross correlation coefficients of Darbandikhan reservoir inflow (I) with different rainfall series (R).

Rainfall Gage Station	Variable	Time Lags					
		t	$t-1$	$t-2$	$t-3$	$t-4$	$t-5$
Darbandikhan	Rd	0.40	0.52	0.52	0.43	0.26	-0.03
Sulaimani	Rs	0.50	0.61	0.60	0.49	0.25	-0.02
Sanadaj	Rn	0.62	0.55	0.45	0.30	0.19	0.01
Kermanshah	Rk	0.50	0.59	0.47	0.31	0.16	0.00
Marivan	Rm	0.57	0.66	0.60	0.44	0.26	-0.02

Table-2: Selected input variables which are used for developing the ANFIS models.

Variable	Selected Time Lags	Time Series Description
I	I_{t-1}; I_{t-2}; I_{t-11}; I_{t-12}	Inflow to Darbandikhan reservoir
Rd	Rd_t; Rd_{t-1}; Rd_{t-2}	Rainfall recorded at Darbandikhan station/Kurdistan Region/Iraq
Rs	Rs_t; Rs_{t-1}; Rs_{t-2}	Rainfall recorded at Sulaimani station/Kurdistan Region/Iraq
Rn	Rn_t; Rn_{t-1}; Rn_{t-2}	Rainfall recorded at Sanadaj station/Iran
Rk	Rk_t; Rk_{t-1}; Rk_{t-2}	Rainfall recorded at Kermanshah station/Iran
Rm	Rm_t; Rm_{t-1}; Rm_{t-2}	Rainfall recorded at Marivan station/Iran

B. Model Structure Development

The selection of appropriate input variables is one of the most critical steps in preparing a satisfactory predicting model as these variables determine the structure of the model and affect the results of the model.

As mentioned earlier, nineteen variables were selected to be used as inputs in the developed ANFIS models. To determine which variables should be the input arguments to produce a best ANFIS model, there are 19 input candidates $(I_{t-1}, I_{t-2}, \dots, Rm_{t-2})$ and the output to be predicted is (I_t). For best model selection, more computationally intensive approach is to do an exhaustive search on all possible combinations of the input candidates. However, the approach usually involves a significant amount of computation if all combinations are tried. For instance, if 4 inputs are selected out of 19, the total number of ANFIS models is $C(19,4) = 19!/((19-4)! \times 4!) = 3876$. To achieve the computation for combinations up to 10 input variables, a MATLAB code have been written that search for best ANFIS model based on the minimum $RMSE$ for checking data set. For example, Figure-5 shows the selection of best model for several inputs combinations (up to 3) according to the minimum $RMSE$ for checking data set.

There are different types of MFs in ANFIS and different number of MFs can be introduced for each input. In the grid partitioning method, ideally for n MFs and p input variables there could be n^p different if-then rules. Consequently, increasing the number of membership functions on the input variables will increase the number of fuzzy if-then rules; simultaneously, it increases the model complexity and hence affects the model parsimony. Moreover, a fuzzy model with a large number of rules often encounters the risk of over fitting the data [8]. In the present study, to obtain the best ANFIS structure, MFs of 2 to 5 for inputs with two different types of MFs (triangular and generalized bel) were tested and trained. No significant improvement in model performance is observed with respect to the change in number of MFs above 5 for triangular MF and above 4 for generalized bel MF. However, as the number of MFs increases, the time taken for model training also increases considerably. Consequently, by the principle of parsimony, the model with 4 triangular MFs and with 3 generalized bel MFs was selected as best numbers of MFs for further analysis. Figure-6 shows the input variables $(I_{t-1}, I_{t-12}, Rm_t)$ for the best models using 4 triangular MFs and 3 generalized bel MFs.

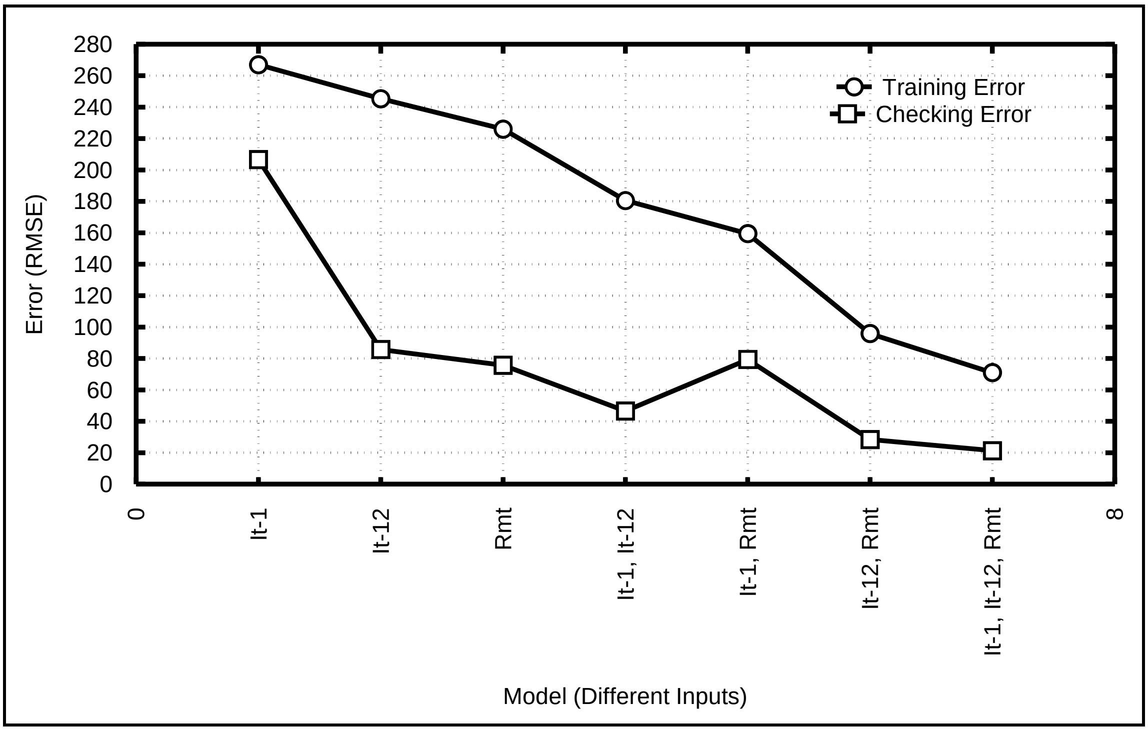

Figure-5: Selection of best model (several inputs combinations) based on the minimum RMSE for checking data set.

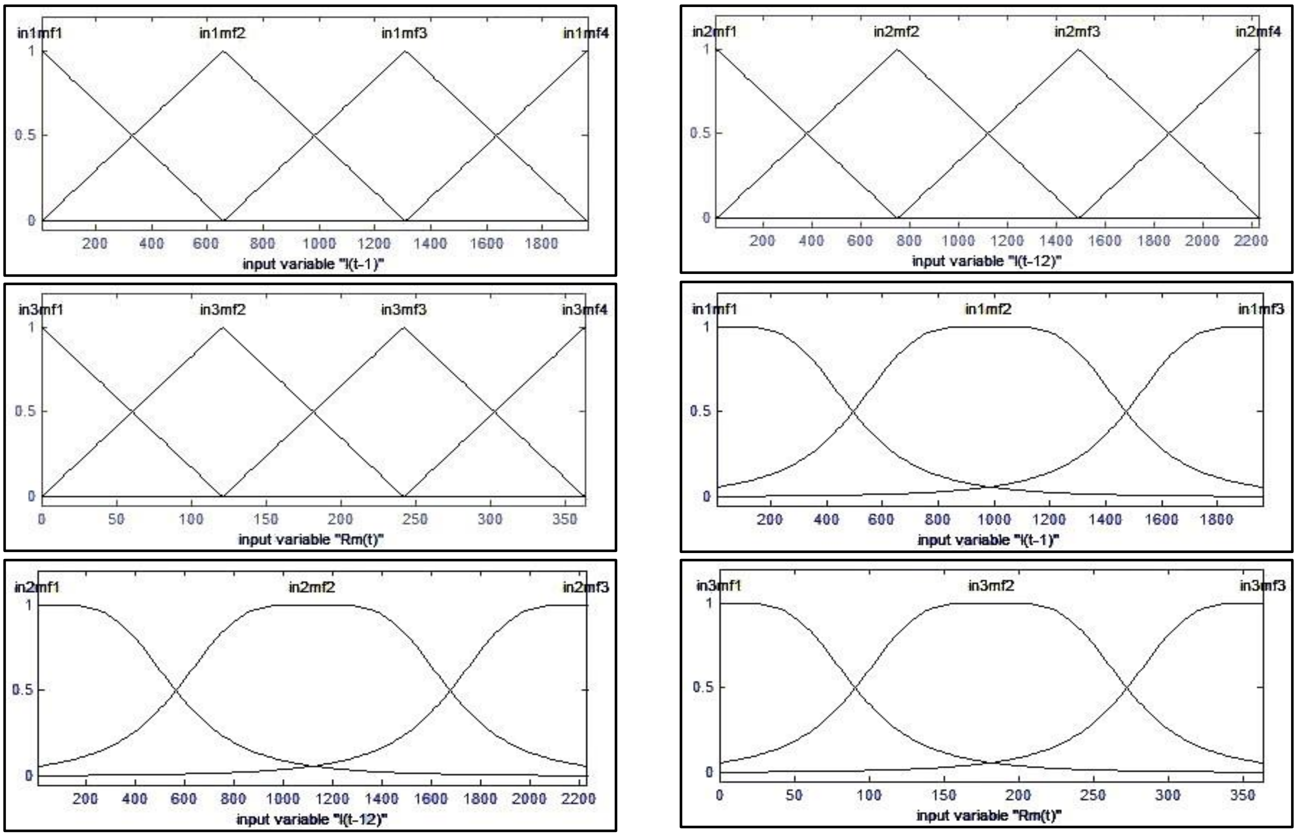

Figure-6: Triangular and generalized bell membership functions for inflow and rainfall input variables for best models.

Results and Discussion

The ANFIS methodology is presented for the purpose of making predictions from past records of reservoir inflow combined with rainfall data. To develop ANFIS model, nineteen variables of reservoir inflow and rainfall of five stations, with their antecedents were selected as inputs to the model. The input variables are four antecedent values at 1, 2, 11, and 12 time lags of reservoir inflow and 0, 1 and 2 time lags of each rainfall of five stations, and the output is the current value of the reservoir inflow. As mentioned earlier, the selection of variables is based on the autocorrelation function for inflow time series and on the cross-correlation function for inflow with different rainfall time series. In ANFIS models, the selections of effective variables with evaluating the correlation coefficient between input and output variables are necessary. Since the number of variables (19 variables) is high, the most effective model variables are found by using a combination of variables, 2 to 5 variables, out of the nineteen variables. On the other hand, two different MFs, triangular and generalized bell, with different numbers, 2 to 5 MFs for each one, were investigated. Therefore, a huge number of ANFIS models were trained by using a MATLAB code had been written for this purpose. To compare the outputs of the models to the observed values and evaluate the applicability of different ANFIS models as well as type of input variables and combinations, RMSE was calculated for checking data set in the code and used to select the best fit model. Twenty two prediction ANFIS models were selected, eleven models for triangular MF and eleven models for generalized bell MF as shown in Table-3 and Table-4 respectively. As the number of input variables and number of MFs are increased, the model structure becomes more complicated. In this study, the grid partition and hybrid algorithm provided by the MATLAB ANFIS toolbox were used and the learning procedure, and the construction of the rules were provided by this algorithm. Furthermore, the performances of the selected ANFIS models were evaluated according to three statistical criteria; coefficient of determination (R^2), coefficient of efficiency (CE) and normalized root mean square error ($NRMSE$).

As seen in Tables 3 and 4, the ANFIS models are evaluated based on their performance in training and checking data sets. The results show, generally, that combining of antecedent inflow with rainfall in the basin as input variables to the model improves the models performance. As it is clear, the most effective model variables among the nineteen variables are the time lags 1, 2, 11 and 12 of inflow (I_{t-1}, I_{t-2}, I_{t-11}, I_{t-12}), time lag 0 of rainfall at Sulaimani station (Rs_t), time lags 0 and 2 of rainfall at Darbandikhan station (Rd_t, Rd_{t-2}), time lags 0 and 2 of rainfall at Kermanshah station (Rk_t, Rk_{t-2}), time lag 2 of rainfall at Sanandaj station (Rn_{t-2}) and time lag 0 rainfall at Marivan station (Rm_t).

The models have shown significant variations in the performance evaluation. Also, comparing the results of the developed ANFIS models, both triangular and generalized bel MFs, approximately, have the same performance with a little difference in the number of MFs. In addition, it is observed that the performance of the M7 and M17 models consisting of three input variables is better than other ANFIS models. Moreover, the best ANFIS models (M7 and M17), both triangular and generalized bel MFs, have the same three input variables (I_{t-1}, I_{t-12}, Rm_t) with the number of MFs, 4 and 3, for triangular and generalized bel MF respectively. The time lags 1 and 12 of the inflow, as shown in the autocorrelation function, have the higher correlation factors with the inflow, therefore, it is convenient to be the most effective input variables to the best ANFIS models.

Table-3: Best models for different inputs based on different numbers of triangular membership function.

Model	Number of Inputs	Number of MF	Inputs Variables (Best Model)	R^2		CE		$NRMSE$	
				Training Data Set	Checking Data Set	Training Data Set	Checking Data Set	Training Data Set	Checking Data Set
M1	2	2	$I_{t-1}; Rs_t$	0.78	0.91	0.78	0.85	0.52	0.34
M2		3	$I_{t-1}; Rm_t$	0.91	0.88	0.91	0.86	0.34	0.33
M3		4	$I_{t-1}; Rm_t$	0.92	0.92	0.92	0.89	0.31	0.29
M4		5	$I_{t-1}; Rm_t$	0.93	0.93	0.93	0.90	0.28	0.28
M5	3	2	$I_{t-1}; I_{t-11}; Rs_t$	0.81	0.92	0.81	0.91	0.48	0.27
M6		3	$I_{t-1}; Rk_{t-2}; Rm_t$	0.92	0.98	0.92	0.97	0.32	0.15
M7		4	$I_{t-1}; I_{t-12}; Rm_t$	0.96	0.95	0.96	0.95	0.23	0.20
M8		5	$I_{t-1}; I_{t-12}; Rm_t$	0.98	0.88	0.98	0.83	0.14	0.37
M9	4	2	$I_{t-1}; I_{t-2}; Rk_{t-2}; Rm_t$	0.93	0.96	0.93	0.96	0.29	0.18
M10		3	$I_{t-1}; Rn_{t-2}; Rk_{t-2}; Rm_t$	0.97	0.88	0.97	0.88	0.18	0.30
M11	5	2	$I_{t-1}; I_{t-2}; Rd_t; Rs_t; Rm_t$	0.98	0.93	0.98	0.87	0.14	0.32

Table-4: Best models for different inputs based on different numbers of generalized bell membership function.

Model	Number of Inputs	Number of MF	Inputs Variables (Best Model)	R^2		CE		$NRMSE$	
				Training Data Set	Checking Data Set	Training Data Set	Checking Data Set	Training Data Set	Checking Data Set
M12	2	2	$I_{t-1}; Rm_t$	0.89	0.90	0.89	0.86	0.37	0.33
M13		3	$I_{t-1}; Rs_t$	0.85	0.91	0.85	0.88	0.43	0.30
M14		4	$I_{t-1}; Rm_t$	0.93	0.89	0.93	0.83	0.28	0.36
M15		5	$I_{t-1}; Rs_t$	0.92	0.90	0.92	0.86	0.32	0.33
M16	3	2	$I_{t-1}; I_{t-12}; Rd_t$	0.89	0.91	0.89	0.91	0.37	0.27
M17		3	$I_{t-1}; I_{t-12}; Rm_t$	0.96	0.96	0.96	0.95	0.21	0.20
M18		4	$I_{t-1}; I_{t-12}; Rm_t$	0.53	0.85	0.06	0.73	1.07	0.45
M19		5	$I_{t-1}; I_{t-2}; Rs_t$	0.97	0.85	0.97	0.75	0.18	0.44
M20	4	2	$I_{t-1}; Rd_{t-2}; Rn_{t-2}; Rm_t$	0.98	0.95	0.98	0.94	0.17	0.21
M21		3	$I_{t-2}; I_{t-11}; I_{t-12}; Rk_t$	0.88	0.73	0.87	0.58	0.40	0.57
M22	5	2	$I_{t-1}; I_{t-2}; Rd_{t-2}; Rk_t; Rm_t$	0.99	0.94	0.99	0.91	0.13	0.26

Furthermore, the reason for the third most effective input variable to be the rainfall recorded at Marivan station in Iran is that the station, approximately, located at the center of the catchment area and close to the main tributary of Darbandikhan reservoir. On the other hand, comparison of results of M7 and M17 models, performance of the M17 model is slightly better than that of M7 model. The best ANFIS model for triangular MF (M7) shows higher R^2 values of (0.96, 0.95), higher CE values of (0.96, 0.95) and $NRMSE$ values of (0.23, 0.20) for the training and checking data sets respectively as shown in Table-3. For the models that use the generalized bell MF, the best ANFIS model (M17) has the highest R^2 values of (0.96, 0.96), highest CE values of (0.96, 0.95) and $NRMSE$ values of (0.21, 0.20) for the training and checking data sets respectively as shown in Table-4. Note that all these models are trained based on 18 months' ahead prediction, which is held out as checking data set.

Further study was carried out on the best ANFIS models, M7 and M17, that use the triangular and generalized bel MFs respectively. The models have been used to predict the inflow to Darbandikhan reservoir, 6, 12, 18 and 24 months ahead (checking data set) as shown in Table-5. The Models have good results, and their statistical evaluation criteria are close together for all predicted periods. With regard to the predict periods, it is evident that the performance of the models in terms of R^2, CE and $NRMSE$ for predicted and observed values of reservoir inflow for checking data set were found to relatively deteriorate at higher lead times. Figure-7 shows graphically the comparison of the predicted versus observed values of the two best ANFIS models, M7 and M17, for Darbandikhan reservoir inflow in different ahead period prediction for the training and checking data sets. The figure illustrates that a clear small underestimation or overestimation exists in the monthly inflow predication value of each model. Avery close fit was obtained between computed and observed inflows up to 18 months ahead for model M17.

Table-5: Performance of the best two models (M7 and M17) using different numbers of checking data set.

Model	Membership Function	Inputs Variables	Checking Data Set (month)	R^2		CE		$NRMSE$	
				Training Data Set	Checking Data Set	Training Data Set	Checking Data Set	Training Data Set	Checking Data Set
M7	Triangular (4 MF)	$I_{t-1}; I_{t-12}; Rm_t$	6	0.96	0.96	0.96	0.93	0.23	0.19
			12	0.96	0.96	0.96	0.95	0.23	0.15
			18	0.96	0.95	0.96	0.95	0.23	0.20
			24	0.97	0.90	0.97	0.82	0.18	0.36
M17	Generalized bell (3 MF)	$I_{t-1}; I_{t-12}; Rm_t$	6	0.97	0.97	0.97	0.96	0.21	0.15
			12	0.97	0.96	0.97	0.94	0.21	0.17
			18	0.96	0.96	0.96	0.95	0.21	0.20
			24	0.97	0.93	0.97	0.91	0.19	0.26

Table-6 shows some descriptive statistics of the observed and predicted inflow time series for training and checking data sets for best two ANFIS models, M7 and M17, which are use 4 triangular and 3 generalized bell MFs respectively. An analysis to assess the potential of each of the models to preserve the statistical properties of the historical inflow series reveals that the inflow series computed by the best ANFIS model with generalized bell MF (M17), generally, reproduces the first two statistical moments (mean and standard deviation) better than that computed by the best ANFIS model with triangular MF (M7) for training and checking data sets. Moreover, generally, the minimum and maximum values of predicted inflow time series for model M17are closer to observed values than that for model M7 for training and checking data sets. The developed ANFIS models can be successfully applied to establish a monthly reservoir inflow prediction model for Darbandikhan reservoir. In this case the future value of input variable (Rm_t) should be predicted. However, it should be kept in mind that such predicting exercises are data dependent. More reliable data will be required to achieve reliable future predicts. Moreover, depending on the data availability more effective parameters such as the temperature, humidity could be considered.

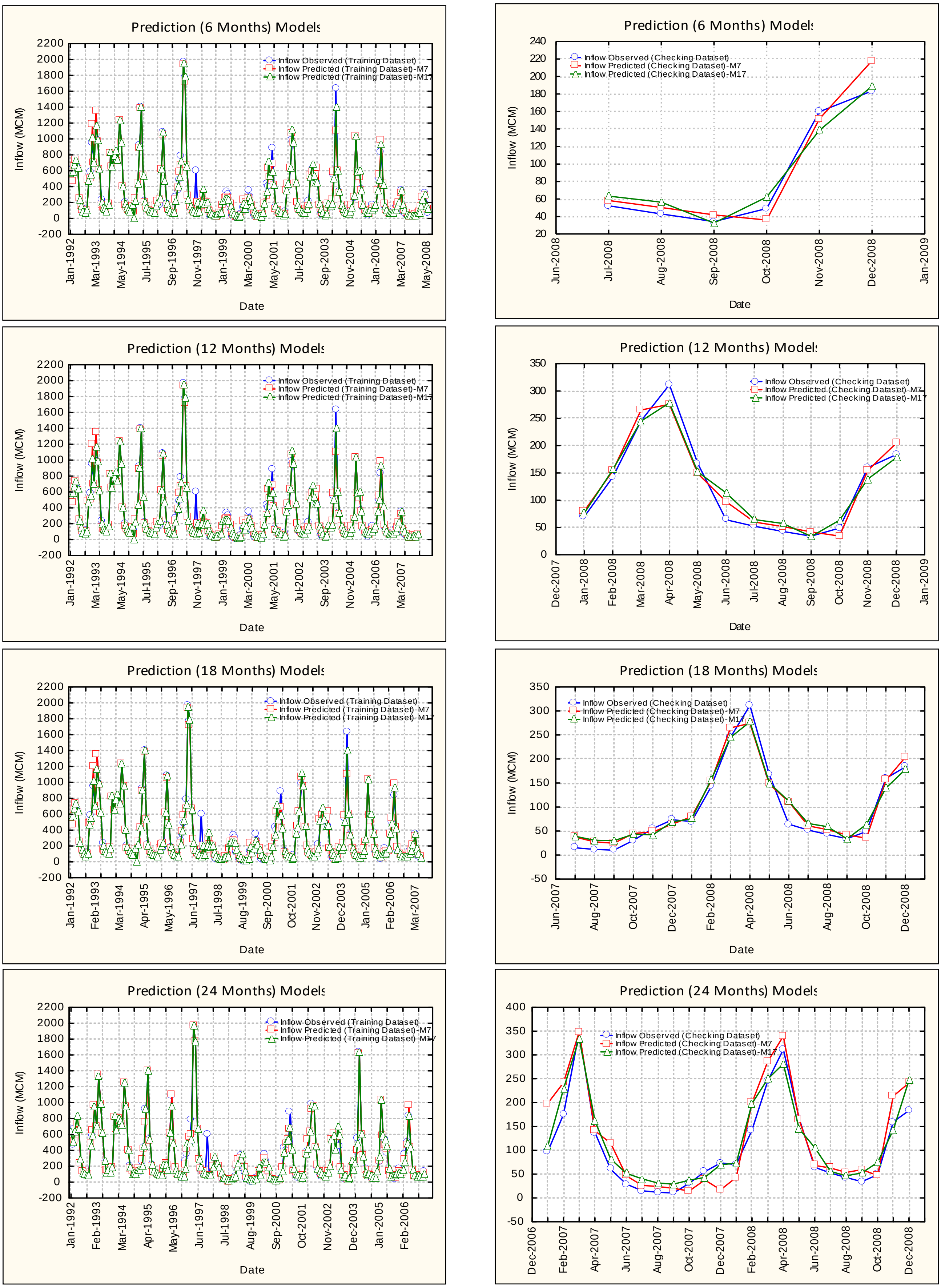

Figure-7: Comparison of observed and predicted values for training and checking data sets for best models (M7 and M17).

Table-6: Descriptive Statistics of observed and predicted inflow for best ANFIS models (M7 and M17).

Variable	Number of Data	Mean	Minimum	Maximum	Standard Deviation
Inflow Observed (Training Data Set)	198	302.75	5.79	1960.73	343.16
Inflow Predicted (Training Data Set)-M7	198	301.78	16.72	1935.07	333.13
Inflow Predicted (Training Data Set)-M17	198	302.48	1.43	1952.52	336.10
Inflow Observed (Checking Data Set)	6	86.93	34.21	183.51	66.29
Inflow Predicted (Checking Data Set)-M7	6	92.65	35.99	218.06	74.77
Inflow Predicted (Checking Data Set)-M17	6	90.30	32.57	188.98	60.09
Inflow Observed (Training Data Set)	192	307.00	5.79	1960.73	347.29
Inflow Predicted (Training Data Set)-M7	192	306.00	18.87	1935.07	337.11
Inflow Predicted (Training Data Set)-M17	192	306.73	2.25	1952.52	340.12
Inflow Observed (Checking Data Set)	12	126.85	34.21	311.65	89.81
Inflow Predicted (Checking Data Set)-M7	12	130.65	34.29	275.64	84.18
Inflow Predicted (Checking Data Set)-M17	12	129.70	33.63	278.28	76.92
Inflow Observed (Training Data Set)	186	315.84	5.79	1960.73	349.26
Inflow Predicted (Training Data Set)-M7	186	314.82	17.63	1935.07	338.85
Inflow Predicted (Training Data Set)-M17	186	315.58	3.01	1952.52	341.91
Inflow Observed (Checking Data Set)	18	95.50	10.37	311.65	86.61
Inflow Predicted (Checking Data Set)-M7	18	101.65	24.67	274.27	80.50
Inflow Predicted (Checking Data Set)-M17	18	100.70	29.97	277.40	75.17
Inflow Observed (Training Data Set)	180	321.66	5.79	1960.73	353.06
Inflow Predicted (Training Data Set)-M7	180	321.75	5.52	1969.53	348.36
Inflow Predicted (Training Data Set)-M17	180	321.68	14.25	1960.67	347.70
Inflow Observed (Checking Data Set)	24	106.92	10.37	347.85	93.68
Inflow Predicted (Checking Data Set)-M7	24	125.42	14.48	348.24	107.85
Inflow Predicted (Checking Data Set)-M17	24	120.43	28.85	334.57	90.52

The results demonstrate that ANFIS can be successfully applied to establish accurate and reliable reservoir inflow prediction models using the values of antecedent inflow and rainfall in the reservoir basin as input variables. Also, it can be observed that, the ANFIS when employed for monthly inflow data prediction, can achieve satisfactory performances for simulating the monthly inflow of Darbandikhan reservoir, and the method has high consistency and good stability. However, both best ANFIS models developed in this study are local models that are limited to applications within the study area and with the observed specific conditions. For other study areas, new ANFIS architectures may be developed using different input variables and lag time relations for predicting monthly reservoir inflow based on the specific characteristics of the basin.

Conclusions

Predicting of reservoir inflow is challenging because of the hydrologic system complexity. Enhancing the quality of reservoir inflow predicting has always been an important task for researchers and hydrologic predictors. To improve the predicting accuracy of inflow to the Darbandikhan dam reservoir which is one of the most water resources in Kurdistan Region, Iraq, this study develops predicting models based on the ANFIS technique to predict the monthly inflow data series. In ANFIS method, there are not any limiting assumptions such as linearity, normality, stationarity, ergodicity, independence of residuals, etc. Moreover, it does need estimations of several order autocorrelation and cross correlation coefficients for prediction and yields more accurate predictions. In addition, the ANFIS algorithm can capture a nonlinear relationship between inputs and outputs, and this is particularly useful in modeling reservoir inflow, where input and output relationships are of a complex and nonlinear nature.

It would be useful to investigate the inclusion of rainfall in reservoir inflow predicting models to improve the predicting accuracy. Therefore, the inclusion of monthly rainfall in monthly reservoir inflow predicting models were carried out using a cross correlation analysis. For this purpose, ANFIS models based on different combinations of monthly data for Darbandikhan reservoir inflow and rainfall of five stations within the basin in Iraq and Iran with their antecedents as input variables were constructed and trained. The data of inflow and rainfall covers the period from January 1992 to June 2007, and from July 2007 to December 2009 are used for training and checking of monthly inflow prediction models, respectively. Among the MFs, the triangular and generalized bell MFs with different numbers (2-5) for each one are investigated. Four statistical criteria are employed to evaluate the performances of various models in order to select the best fit models. These criteria are the coefficient of determination (R^2), the coefficient of efficiency (CE), the root mean squared error ($RMSE$) and the normalized root mean squared error ($NRMSE$).

The results showed that the predict ANFIS models with input variables such as rainfall in addition to antecedent inflow are effective in 6, 12, 18 and 24 months ahead predicting of Darbandikhan reservoir inflow. The models have shown better generalization capability and ability to improve the reservoir inflow prediction accuracy. Also, the results of predicting monthly inflow to the reservoir indicated that the model uses a combination of antecedent inflows (I_{t-1}, I_{t-12}) and rainfall at Marivan station in Iran (Rm_t) as input variables displayed better performance, in terms of statistical evaluation criteria, than other models. In addition, comparison of triangular MF with generalized bell MF showed that the accuracy of models, approximately, is the same. Furthermore, in general, the results demonstrate that the ANFIS method is an efficient tool and can be successfully applied to establish accurate and reliable reservoir inflow prediction models when there are enough observed data. The results from this study may provide a motivation for monthly reservoir inflow predicting by using a combined inflow with various operationally available climatic variables as input variables to ANFIS model. The developed models have satisfactory performances in monthly inflow prediction of Darbandikhan reservoir and will be helpful for reservoir managers to obtain more accurate and stable predicting results. The findings of this study may encourage the directorate of Darbandikhan dam to predict the reservoir inflow by using ANFIS method with various operationally available climatic variables as input variables. Temperature, humidity, …, etc. can be combined into inflow and rainfall for possible predict improvement.

References

[1] A. SHALAMU, *Ph.D Thesis,* Las Cruces, New Mexico: New Mexico State University, 2009.

[2] C.-T. Cheng, Z.-K. Feng, W.-J. Niu and S.-L. Liao, "Heuristic Methods for Reservoir Monthly Inflow Forecasting: A Case Study of Xinfengjiang Reservoir in Pearl River, China," *Water,* vol. 7, pp. 4477-4495, 2015.

[3] H. Galavi and L. T. Shui, "Neuro-Fuzzy Modelling and Forecasting in Water Resources," *Scientific Research and Essays,* vol. 7, no. 24, pp. 2112-2121, 2012.

[4] A. ALTUNKAYNAK, M. OZGER and M. CAKMAKCI, "Water Consumption Prediction of Istanbul City by Using Fuzzy Logic Approach," *Water Resources Management,* vol. 19, no. 5, pp. 641-654, 2005.

[5] M. Firat, M. E. Turan and M. A. Yurdusev, "Comparative analysis of Fuzzy Inference Systems for Water Consumption Prediction," *Hydrology,* vol. 374, no. 3, pp. 235-241, 2009.

[6] M. T. Dastorani, H. Afkhami, H. Sharifidarani and M. Dastorani, "Application of ANN and ANFIS Models on Dryland Precipitation Prediction (Case Study: Yazd in Central Iran)," *Applied Sciences,* vol. 10, no. 20, pp. 2387-2394, 2010.

[7] J.-Y. Lin, C.-T. Cheng, Y.-G. Sun and K. Chau, "Long-Term Prediction of Discharges in Manwan Hydropower Using Adaptive-Network-Based Fuzzy Inference Systems Models," *Lecture Notes in Computer Science,* vol. 3612, pp. 1152-1161, 2005.

[8] P. C. Nayak, K. P. Sudheer, D. M. Rangan and K. S. Ramasastri, "Short-term Flood Forecasting With a Neurofuzzy Model," *Water Resources Research,* vol. 41, no. W04004, pp. 1-16, 2005.

[9] K. Shahverdi, K. Khodaei and A. Shokohi, "Prediction of Monthly Inflow to Dam Reservoir Using ANN and ANFIS: A Case Study of Latyan Dam," *International Bulletin of Water Resources & Development,* vol. III, no. 03, pp. X-XVIII, 2015.

[10] “Darbandikhan Basin Project,” Nature Iraq/Twin Rivers Environmental Research Institute/The American University of Iraq-Sulaimani, Sulaimani, Kurdistan, Iraq, 2008.

[11] “Darbandikhan Lake Poisoning Event,” Nature Iraq/Twin Rivers Environmental Research Institute/The American University of Iraq-Sulaimani, Sulaimani, Kurdistan, Iraq, 2008.

[12] M. C. Cordell, “Consultancy Services for Dokan and Derbendikhan Dam Inspections,” The World Bank/SMEC International Pty. Ltd., Melbourne, Australia, 2006.

[13] “Islamic Republic of Iran Meteorological Organization,” Chaharmahal va Bakhtiari Meteorological Administration, [Online]. Available: http://www.chaharmahalmet.ir/iranarchive.asp. [Accessed 21 8 2016].

[14] S. Makridakis, S. C. Wheelwright and R. J. Hyndman, Forecasting, Methods and Applications, 3rd ed., New York: John Wiley and Sons, Inc., 1998.

[15] A. GEORGE and U. C. IOANA, “Forecasting Tourism Demand Using ANFIS for Assuaring Successful Strategies in the View of Sustainable Development in the Tourism Sector,” in *2nd IASME / WSEAS International Conference on Energy & Environment*, Portoroz, Slovenia, 2007.

[16] A. K. Palit and D. Popovic, Computational Intelligence in Time Series Forecasting, Theory and Engineering Applications, London: Springer-Verlag London Limited, 2005.