Research Article

Comparison of the hydrological time series modeling by the floods in river Indus of Pakistan

Salman Bin Sami,[1] Sobia Shakeel,[2] Reema Salman[3]
[1]Management Science Department, SZABIST (Karachi Campus), Pakistan
[2]Management Science Department, SZABIST 100 (Karachi Campus) Block 5 Clifton, Pakistan
[3]Department of Mathematics, University of Karachi, Pakistan

Correspondence: Mr. Salman Bin Sami, Management Science Department, SZABIST (Karachi Campus), Karachi, Pakistan, Tel +92-3333175611, Email salmankhan909@live.com

Received: June 01, 2022 | **Published:** July 21, 2022

Abstract

Today, in the field of science and technology, huge forecasting applications are used by scholars to forecast future values. Nowadays, using estimating the flood forecasting for peak flow discharges is very common for the risk assessment annually by quantitative data collections from different resources. The very famous and longest rivers of Pakistan i.e. Indus River and other rivers too like River Jhelum, River Kabul, and River Chenab are the prime sources of flooding. These rivers are the prime tributaries of the Indus River System. Pakistan's longest river, River Indus, is connected with the seven (7) gauge stations called Dams and barrages, and they are playing a vital role in the generation of electricity and also in irrigation for Pakistan. In this research paper, we calculated the flood risk for the Indus using the streamflow discharges on the daily basis. At present, Adaptive Neuro-Fuzzy Inference System (ANFIS) model is widely used to analyze these hydrological time series data. Adaptive Neuro-Fuzzy Inference Systems (ANFIS) merges the potentiality of Fuzzy Inference Systems (FIS) and Artificial Neural Networks (ANN) to work out problems of different kinds. For this purpose, we used the data for the years from 2002 to 2012 daily (6-months each year) streamflow period. In our analysis, the root means square error (RMSE) shows that the ANFIS model generated more satisfactory results than other models with minimum prediction errors. The ANFIS model is more reliable and has the feasibility of integrating the essence of a fuzzy system into the real world.[1–28]

Keywords: neuro-fuzzy network, fuzzy logic, fuzzy inference system, hydrological modeling, river Indus, adaptive neuro-fuzzy inference systems

Introduction

Nowadays, floods are a serious alarm for everyday life occurring wherever naturally in the world. Humans, animals, and all living things are influenced by the damage from the flood. It will damage not only our lives but destroys vegetation and the environment too. During floods, we can face big losses sustained to the economy. Just take a look at the past few decades for the flood experiencing countries, due to global warming situations a rapidly growing in floods.[29] Therefore, all scientists, economists, engineers, etc. daily plan to get accurate time or want to know before the time of these events.

Ologunorisa and Abawa, in 2005, explained various methods like hydrological equipment, barometrical tools and conditions, socioeconomic elements, and a combination of hydro-barometrical and socioeconomic elements along with a geological data network to estimate flood risk.[28]

Smith[22] analyzed that the probability of the occurring events along with its results is also essential to be observed for estimating flood risk. Variability in water resources leads to heterogeneity in geological expansions naturally and variance in complex socio-economic features. Khan et al.[30] utilized historical compiled data of the highest peak discharges in Pakistan. He evaluated the flood risk of the Indus River by assessing the probabilities of the occurrence of a flood.

Several studies reveal that countries such as Pakistan, Korea, the USA, and many others widely employ barometrical parameters for flood risk assessment. Kalma and Laughlin[11] employed an approach based on local weather data along with an area to graph or sketch flood risk. A researcher, Khan studied the risks of flood in the affected neighborhood regions of the Indus River in Pakistan. He applied a better and more efficient technique named the GIS technique. It required digital image processing, a geological data system, and remote sensing.[29] He utilized satellite data that emphasized that it is

significant to make dams to reduce flood risks. A similar procedure was adopted by Nawaz and Shafique in 2003[26] on the river Jhelum. Various forecasting approaches for Rivers and Dams were attempted fortunately using the Linear and non-linear regression techniques by Burn and McBean in 1985,[7] Awwad, El-Fandy, and Karunanithi in 1994.[16,18,19] All these researches gave a better forecast of dams' river flow.

Selas and Smith used hydrological time series modeling to develop synthetic stream flows.[5] Similarly, Stedinger and Taylor generated five various models by assuming stream flow images.[6] Researchers are widely using time series forecasting in different fields like physics, engineering, medicine, and finance as well. The consistently used modeling and forecasting approaches by researchers for time series are AR (Auto-regressive) method formulated in 1970,[1] ARMA (Auto-regressive moving average), and ARIMA (Auto-regressive integrated moving average) disaggregation models developed by Valencia and Schaake in 1973,[2] and many other.

Time series forecasting is considered the process that forecasts weather using time series data. These methods are applicable for time series data only for making forecasts. This time series data, utilized for months of massive stream flows such as many hours or even a day, is in the arithmetic data form.

Hassan and Ansari, in 2010, forecasted the continuous behavior of River Indus by employing various nonlinear methods.[32] Sudheer used an ANN model for a similar goal and worked out that the ANN model required further advancements for forming the peak data flows accurately.[24]

We have studied the ANFIS (adaptive neuro-fuzzy inference system) model for the Indus basin in our research. Similarly, another investigation was done by Nayaka and Sudheer, in India, in which they employed the ANFIS model for evaluating a hydrological model of time series for Baitarani River's basin stream flow in Orissa state.[24]

Various complex hydrologic modeling systems have highly systematic tools for forecasting. These include genetic algorithms, adaptive neuro-fuzzy inference systems (ANFIS), and artificial neural networks (ANN). In 1965, the Fuzzy logic approach was developed so that a decision-making and expertise system similar to humans can be described.

Recently, Tayyab et al., 2018, have compared two decomposition-based models, ensemble empirical mode decomposition (EEMD) and discrete wavelet transform (DWT) with an artificial intelligence-based model to forecast streamflow at the upper Indus basin.[36] Results indicate that the decomposition-based models gave better prediction accuracy, especially ensemble empirical mode decomposition outperforming all the models. Nazir et al., in 2019, employed Variational Mode Decomposition (VMD) model that is based on a denoising technique called singular spectrum analysis (SSA), Empirical Bayes Threshold (EBT), and Support Vector Machine (SVM).[37] They applied these models to predict the daily river inflow of the Indus River Basin and compared the proposed model with others. Results showed that the suggested gave superior results and is validated for power-generating systems and water resources management.

The main significance of the ANFIS model is that it can maintain the full capacity of the ANN method along with the method's simplicity. In 1993, Tagaki-Sugeno-Kang (TSK) and Yasukawain formulated ANFIS which is the mapping of the fuzzy or fuzzy-rule-based algorithms.[8] Since the last decade, scientists are using ANFIS widely for water resources predictions. Today, numerous applications, like the prediction of water resources and planning and database management, are also using ANFIS.

This paper is organized into six sections. Section 2 defines our study area. Section 3 evaluates the materials and methods employed in this research. Section 4 provides the performance evaluation of the models. Section 5 presents results and discussion and Section 6 concludes this research.

Study area

To study the severity of damages by floods we used extensively technologies constructively or non-constructively. The constructive approach requires measuring a large amount of time and money as well. This approach includes some facts about making dams and reservoirs and also changing the flow of rivers. On the other hand, the non-constructive measures are dealing with relief when floods occur and planning for the forecasting of floods to provide such services to the victims. Here, for the prediction of future values with the help of past data the time series analysis forecast is used.

Pakistan is an Asian country that is in the Western zone of this subcontinent. It lies in the 23-37 degrees in the north and from 60-77 degrees in the east. It is comprised of five provinces, namely Sindh, Punjab, Gilgit-Baltistan, Balochistan, and Khyber Pakhtunkhwa, along with a tribal region as well. The weather conditions are different with variations in temperature in all these provinces.

Some regions face extreme weather like heavy rain that causes floods. One reason leading to canal floods is the melting of snow on mountains. The history of Pakistan is full of floods among which the floods in 1950, 1956, 1973, 1976, 1978, 1988, 1992, and 2010 were coped. There is variation in the graphs of floods caused from 1922 to 2010. Among all, the most disastrous, catastrophic, and unfortunate was the flood that occurred in 2010.

The effects of floods on the defined regions of the Indus River of Pakistan are analyzed in this research. In Figure1, the altitude of the surrounding regions of the flooded areas of Pakistan is shown.[34] Heavy rainfall during the monsoon rains in Pakistan is accompanied by the melting of snow in canals. This leads to calamitous floods. Another significant reason for floods is land sliding. These disastrous floods bring various losses and damages. Few of them include losing the lives of animals and men, huge constructional losses, decomposition of agricultural land, scarcity, and an increase in water transport diseases.

Figure 1 Flooded areas of Pakistan.

In 2010, FFC (Federal Flood Commission) elucidated in their annual report that the streambed of River Indus in Sindh with the neighborhood having peak flow faced the maximum damages.[31]

Among the extended rivers of the world, the Indus River with 1800 miles in length and seven barrages is considered the longest one. 450000 square miles approximately is the aggregate discharged region of this river. Among which 275000 square miles are in the desertification areas and the remaining is in the mountainous regions of Pakistan.

In Pakistan, River Indus runs in the southern direction starting from Ladakh in Jammu Kashmir and finally linking up with the Arabian Sea in Sindh. Figure 2 indicates the seven gauge stations that monitor the River Indus.[35] They are Chashma Barrage, Tarbela Dam, Taunsa Barrage, Jinnah or Kalabagh Dam, Sukkur Barrage, Kotri Barrage, and Guddu Barrage.
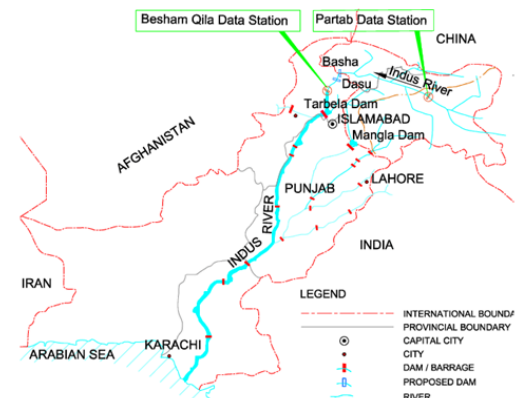

Figure 2 Map conveying core plans of River Indus in Pakistan.

These gauge stations record various levels of flood risk that range from medium to extremely high. Medium level flood risks are noticed at Kalabagh and Tarbela Dams. Taunsa and Chashma Barrages are observed to have high-level flood risks. Whereas, Guddu and Kotri

Barrages fluctuate from high to extremely high risks. Sukkur Barrage notices extremely high risks.

This research analysis suggests using the ANFIS approach to establish a time data series model of river flow for the Basin of River Indus in Pakistan. For this purpose, we utilize annual flood peak discharges via various gauge sites.

Materials and methods

This section gives details on time series modeling along with forecasting. To show the fundamental structure of the series it is significant to distinguish and adopt an appropriate model. Therefore, a tailored model can provide planned future forecasting. The Time series model recognizes the relationship between the current value and the previous observation. Thus, it studies the linear or non-linear values and suggests whether it has a sequence or relationship among the values or not. Nevertheless, various forms of time series models show different stochastic methods as well. Among them Moving Average (MA) and Autoregressive (AR),[6,12,23] are outstanding linear time series models. We suggest a blend of both names, Autoregressive Moving Average (ARMA),[6,12,21,23] and Autoregressive Integrated Moving Average (ARIMA) in this research. Contrastingly, Autoregressive Fractionally Integrated Moving Average (ARFIMA),[9,17] is another model that derives ARMA and ARIMA models. For seasonal time series forecasting, one can use a distinct version of ARIMA which is the Seasonal Autoregressive Integrated Moving Average (SARIMA),[3,6,23] model. All the modified versions of the ARIMA model are widely named the Box-Jenkins Models because the Box-Jenkins principle,[1,8,12,23] drives them. The ability and simplicity of all the linear models to understand and apply are giving them remarkable attention and popularity.

Unfortunately, in various cases, time series give non-linear patterns. To evaluate volatility in financial and economic time series it is appropriate to use non-linear models. According to this, some widely applicable non-linear models include Autoregressive Conditional Heteroskedasticity (ARCH) with its altered versions namely; Generalized ARCH (GARCH), Exponential Generalized ARCH (EGARCH),[9] the Nonlinear Moving Average (NMA)[28] model, the Non-linear Autoregressive (NAR)[7] model, the Threshold Autoregressive (TAR),[8,10] model, and others.

Autoregressive integrated moving average (ARIMA) models

All ARMA models use stationary time series data. However, various time series show non-stationary behavior especially those for business and socio-economic.[23] Time series that possess specific or seasonal patterns also indicate non-stationary behavior.[3,11] As ARMA models fail to evaluate widely applicable non-stationary time series so the ARIMA model,[6,23,27] is suggested.

ARIMA models convert the non-stationary time series into stationary time series via finite differencing of data points. The mathematical representation of ARIMA (p, d, q) with lag polynomials is as follows:[23,27]

$$\varphi(L)(1-L)^d y_t = \theta(L)\varepsilon_t,$$

i.e.

$$\left(1 - \sum_{i=1}^{p} \varphi_i L^i\right)(1-L)^d y_t = \left(1 + \sum_{j=1}^{q} \theta_j L^j\right)\varepsilon_t \qquad (1)$$

I. The integers p, d, and q have a value of zero or greater than zero.

These integers mention the order of integrated, autoregressive, and moving average parts of the model.

II. The level of difference is controlled by the integer d which usually has a value of one. The model contracts to ARMA (p, q) model if d becomes equal to zero.

III. If q = 0 then, ARIMA (p, 0, 0) becomes AR(p) model. Similarly, if p = 0 then, ARIMA (0, 0, q) becomes MA (q) model.

Random Walk model,[8,12,21] is a special case of ARIMA (0, 1, 0) that gives $y_t = y_{t-1} + \varepsilon_t$. This model is commonly applied for non-stationary data, especially in stock price series and economics.

The Autoregressive Fractionally Integrated Moving Average (ARFIMA) model is a practical inference of ARIMA models. The ARFIMA model permits non-integer values of the differencing parameter d. To model time series having long memory,[17] ARFIMA plays a significant role. To expand the term $(1-L)^d$ general binomial theorem is applied. The contributions of various researchers proved to be significant for estimating parameters of general AFRIMA.

Adaptive neural-based fuzzy inference system (ANFIS)

For fuzzy inference that is constructed from fuzzy logic methods and ANNs, we use the ANFIS model that is formulated by Sugeno.[15] It uses a cross-learning rule to identify various parameters. This rule amalgamates the back-propagation gradient descent and the least square method. By applying correct membership functions, ANFIS can serve as a base to build numbers of IF-THEN rules in Fuzzy for producing prior specified input and output pairs Figure 3.[23]
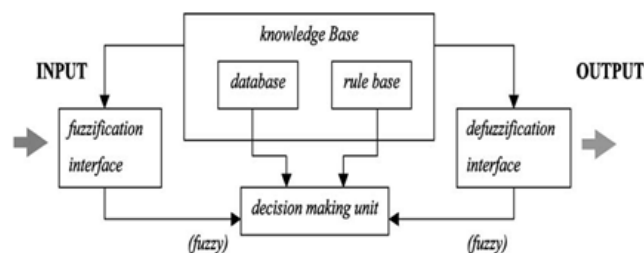

Figure 3 Brisky output by Fuzzy Inference System.

As the inference system of Sugeno fuzzy is mathematically efficient, it can be applied for adaptive, linear as well as optimization techniques. Consider a fuzzy inference of *x* and *y* as two inputs and z as one output in the 1st order. Below is the widely applied rule along with two fuzzy if-then rules:

Rule 1: If *x* is A$_1$ and *y* is B$_1$ then $f_1 = p_1 x + q_1 y + r_1$

Rule 2: If *x* is A$_2$ and *y* is B$_2$ then $f_2 = p_2 x + q_2 y + r_2$

Figure 4a represents the clear outcomes. This figure evaluates the inference system of fuzzy reasoning which is giving *(f)* as an output function while using [*x, y*] as the input vector. The corresponding equivalent ANFIS architecture is a five-layer feed-forward network that is using neural network learning algorithms. These neural network learning algorithms are coupled with fuzzy reasoning for mapping an input space to an output space. This is evident from Figure 4b. The literature has more details and presentations of ANFIS for forecasting hydrological time series.[23,27,33]

We have Sugeno-type, linear combinations of end parameters and their overall output in the proposed model. So, in Figure 3 the output *(f)* can be improved as:

$$f = \overline{w_1}f_1 + \overline{w_2}f_2 = (\overline{w_1}x)p_1 + (\overline{w_1}y)q_1 + (\overline{w_1})r_1 + (\overline{w_2}x)p_2 + (\overline{w_2}y)q_2 + (\overline{w_2})r_2 \qquad (2)$$

The least-squares method computes the end parameters $(p_1, q_1, r_1, p_2, q_2, r_2)$. Thereby, it becomes easier to project the best parameters of the ANFIS model by using a hybrid learning algorithm. To have further explanations one can cite the work by Jang and Sun.[14]
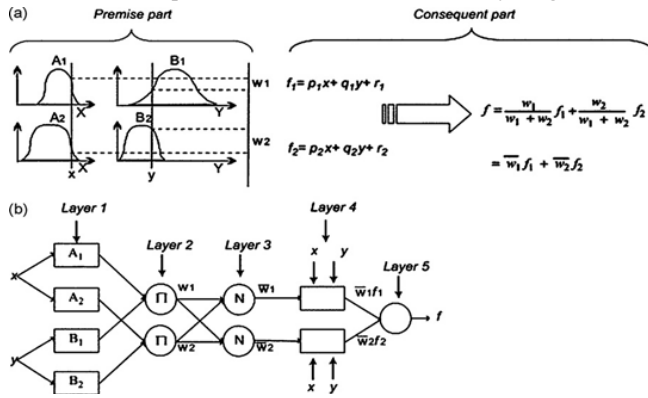

Figure 4 (a) Shows the inference system of Fuzzy. (b) Shows an Equivalent architecture of ANFIS.

Data used

In this paper, we have used the data collected from the source of the Federal Flood Commission (FFC), which is situated in Islamabad-Pakistan, comprised of 11 years recorded by the three gauge stations, Tarbela Dam, Chashma, and Sukkur Barrages, situated at different places.

Performance evaluation of the models

Various researches on the application, validation and calibration of hydrological models recommend only a few approaches to hydrological time series. To evaluate the performance we compute four criteria as stated in the next section.

Different classifications of traditional statistics are regarded as statistical work explanations. For this estimation test, we applied root mean square error (RMSE) that is given as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(d_i^o - d_i^p)^2}{n}} \qquad (3)$$

Here, at any time *t*, the observed flow of the stream is denoted by d_i^o and the predicted flow of the stream is denoted by d_i^p.

Results & discussions

Outcomes and analysis from ANFIS

Investigation of the data reveals to us that the data we surveyed is disordered and highly varied. We can view the behavior of the flow in Tarbela Dam, Chashma, and Sukkur Barrages, from the timeline 2010 to 2012 in Figure 5.

Parameters are provided in Table 1 which are interrelated to these three stations. In Table 1, we can see that the difference between the maxima amount and the minima amount with the standard deviation is very large, therefore the modeling will be complex for the gauge stations. We notice that among the three stations Sukkur Barrage observes the highest maximum amount of peak discharge which is 1130995 fps stream flow for the period 2002-2011, so the range of flood risk is extremely high on it. Table 1 also shows that the least ratio of average to standard deviation is 0.94 fps which is given by Sukkur Barrage.
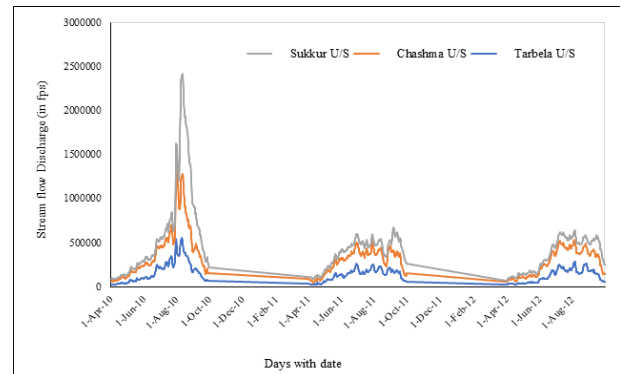

Figure 5 From 2010 – 2012-Stream flows of Chashma Barrage, Sukkur Barrage, and Tarbela Dam.

For the fuzzy-logy network, we are considering input data for ten years (2002 to 2011) for the daily stream flow applying as training data on different models. Also, we tested different models by using the testing data for the year 2012 only and taking each gauge station's stream flows daily data for 6 peak months'. The results in the tables shown below are as entered Input data into the neuro-fuzzy network.

Finding the results, we used RMSE to figure out the outputs as we can see the Tables 2 to 4 are the calculated results. For better outcomes, we used Gaussian membership functions better than the Triangular membership functions for good outcomes with the 0.001 error tolerance.

Tarbela dam results

Outcomes of the Tarbela dam can be viewed in Tables 2a and 2b respectively. The Curves shown below in Figure 6a highlighted the predicted values and surface area using the applied neuro-fuzzy technique for the year 2012 only. Figure 6a shows in the predicted values, on *the y*-axis, the output data means stream values of data in *fps*, and on the *x*-axis, the index means the no. of days for the year 2012 which is 183 peak days. The Curves shown below in Figure 6b highlighted the predicted values and surface area using the applied neuro-fuzzy technique for the year 2002 to 2011. From Figure 6b, we are showing the predicted values. On the *y*-axis, the output data means stream values of data in *fps*, and on the *x*-axis, the index means the no. of days for the mentioned years which is 1830 peak days.

Chashma barrage results

Similarly for Chashma Barrage, as we have done calculations for Tarbela Dam above, tables 3a and 3b are showing the results of the daily stream flow prediction for the year 2012 and from 2002 to 2011. The Curves shown below in Figure 7a highlighted the predicted values and surface area using the applied neuro-fuzzy technique for the year 2012 only. The Curves shown below in Figure 7b highlighted the predicted values and surface area using the applied neuro-fuzzy technique for the year 2002 to 2011.

Sukkur barrage results

Now the outcomes for the Sukkur barrage for taking different inputs and MFs by the daily stream flow as shown below in the following Tables 4a and 4b respectively same as the above calculation techniques. The Curves shown below in Figure 8a highlighted the predicted values and surface area using the applied neuro-fuzzy technique for the year 2012 only. The Curves shown below in Figure 8b highlighted the predicted values and surface area using the applied neuro-fuzzy technique for the year 2002 to 2011.

Table 1 For 10 years & 1 year alone (2002–2011 & 2012) Estimated parameters

Estimated parameters	Tarbela Dam		Chashma Barrage		Sukkur Barrage	
	2002–2011	2012	2002–2011	2012	2002–2011	2012
Average (fps)	141643.2	123941	177492.7	158845.2	116962.8	81611.91
Standard deviation (fps)	88743.25	76086.45	101397.2	79063.51	124973.6	47645.55
Minimum amount (fps)	18800	26000	23493	26169	16405	15630
Maximum amount (fps)	557100	284000	957309	276745	1130995	214780
The ratio of average to standard deviation (fps)	1.6	1.63	1.75	2.01	0.94	1.71

Table 2a 2012 error evaluation by daily flow prediction as testing data

Serial No.	Different input variations	No. of membership functions	RMSE
1	d_t, d_{t-1}	2	113.787
2	d_t, d_{t-1}	3	108.474
3	d_t, d_{t-1}	4	105.902
4	d_t, d_{t-1}	6	104.87
5	d_t, d_{t-1}	8	97.42
6	d_t, d_{t-1}, d_{t-2}	2	101.913
7	d_t, d_{t-1}, d_{t-2}	3	90.465
8	d_t, d_{t-1}, d_{t-2}	4	85.161
9	d_t, d_{t-1}, d_{t-2}	6	68.037
10	d_t, d_{t-1}, d_{t-2}	8	51.114

Table 2b 2002 - 2011 error evaluation by daily flow prediction as training data

| Serial No.	Different input variations	No. of membership functions	RMSE
1	d_t, d_{t-1}	2	155.639
2	d_t, d_{t-1}	3	134.878
3	d_t, d_{t-1}	4	129.603
4	d_t, d_{t-1}	6	125.749
5	d_t, d_{t-1}	8	124.374
6	d_t, d_{t-1}, d_{t-2}	2	129.476
7	d_t, d_{t-1}, d_{t-2}	3	117.745
8	d_t, d_{t-1}, d_{t-2}	4	114.935
9	d_t, d_{t-1}, d_{t-2}	6	110.249
10	d_t, d_{t-1}, d_{t-2}	8	107.461

Table 3a 2012 error evaluation by daily flow prediction as testing data

| Serial No.	Different input variations	No. of membership functions	RMSE
1	d_t, d_{t-1}	2	117.188
2	d_t, d_{t-1}	3	109.666
3	d_t, d_{t-1}	4	107.837
4	d_t, d_{t-1}	6	101.392
5	d_t, d_{t-1}	8	95.455
6	d_t, d_{t-1}, d_{t-2}	2	114.15
7	d_t, d_{t-1}, d_{t-2}	3	100.935
8	d_t, d_{t-1}, d_{t-2}	4	97.08
9	d_t, d_{t-1}, d_{t-2}	6	77.261
10	d_t, d_{t-1}, d_{t-2}	8	69.699

Table 3b 2002 - 2011 error evaluation by daily flow prediction as training data

Serial No.	Different input variations	No. of membership functions	RMSE
1	d_t, d_{t-1}	2	186.708
2	d_t, d_{t-1}	3	169.368
3	d_t, d_{t-1}	4	162.853
4	d_t, d_{t-1}	6	156.526
5	d_t, d_{t-1}	8	156.105
6	d_t, d_{t-1}, d_{t-2}	2	158.58
7	d_t, d_{t-1}, d_{t-2}	3	155.167
8	d_t, d_{t-1}, d_{t-2}	4	148.86
9	d_t, d_{t-1}, d_{t-2}	6	141.91
10	d_t, d_{t-1}, d_{t-2}	8	139.753

Table 4a 2012 error evaluation by daily flow prediction as testing data

Serial No.	Different input variations	No. of membership functions	RMSE
1	dt, d_{t-1}	2	86.433
2	dt, d_{t-1}	3	80.395
3	dt, d_{t-1}	4	75.492
4	dt, d_{t-1}	6	68.732
5	dt, d_{t-1}	8	63.513
6	dt, d_{t-1}, d_{t-2}	2	77.205
7	dt, d_{t-1}, d_{t-2}	3	69.544
8	dt, d_{t-1}, d_{t-2}	4	60.337
9	dt, d_{t-1}, d_{t-2}	6	48.093
10	dt, d_{t-1}, d_{t-2}	8	44.276

Table 4b 2002 - 2011 error evaluation by daily flow prediction as training data

Serial No.	Different input variations	No. of membership functions	RMSE
1	dt, d_{t-1}	2	170.088
2	dt, d_{t-1}	3	149.33
3	dt, d_{t-1}	4	138.725
4	dt, d_{t-1}	6	118.698
5	dt, d_{t-1}	8	112.94
6	dt, d_{t-1}, d_{t-2}	2	135.046
7	dt, d_{t-1}, d_{t-2}	3	114.976
8	dt, d_{t-1}, d_{t-2}	4	106.599
9	dt, d_{t-1}, d_{t-2}	6	93.434
10	dt, d_{t-1}, d_{t-2}	8	87.921

Predicted values
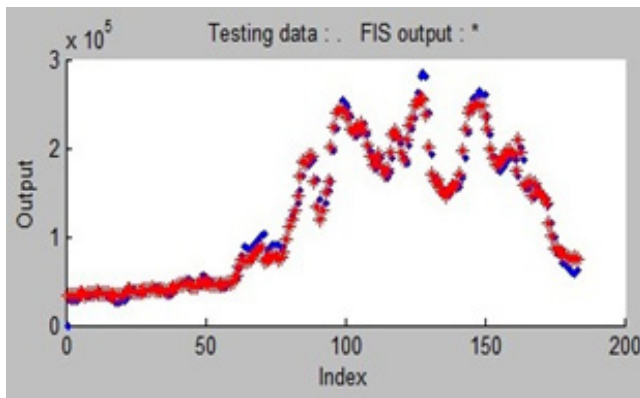
Predicted values
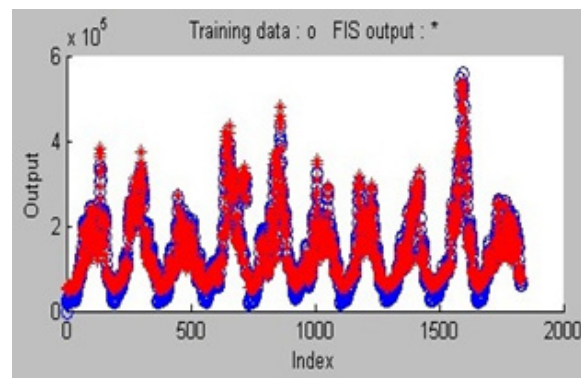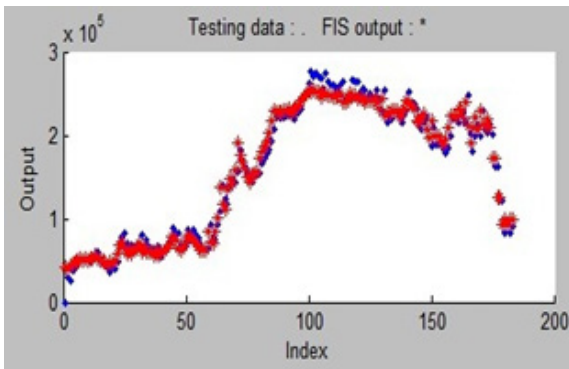

Surface area

Surface area
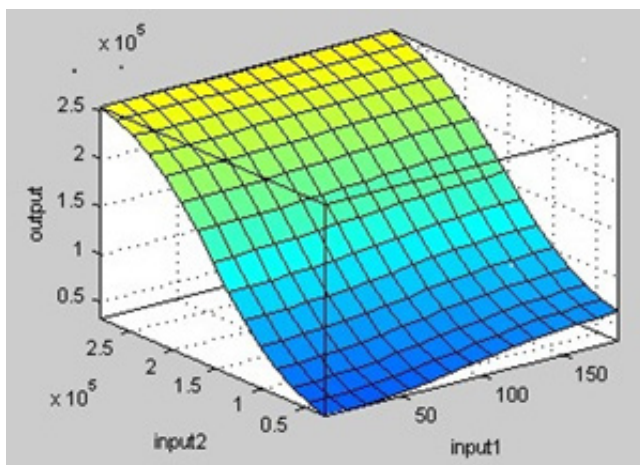
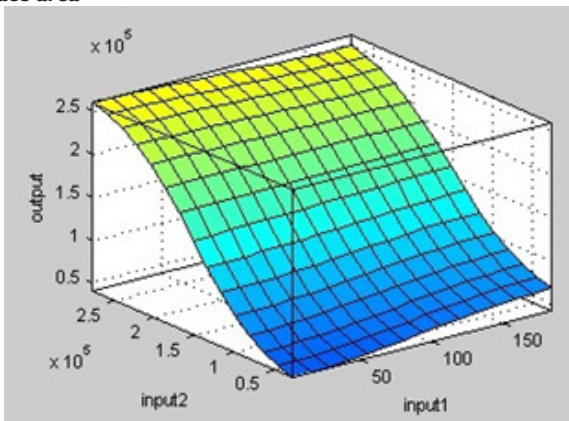**Figure 6b** For the years 2002 to 2011- Predicted values and surface area.
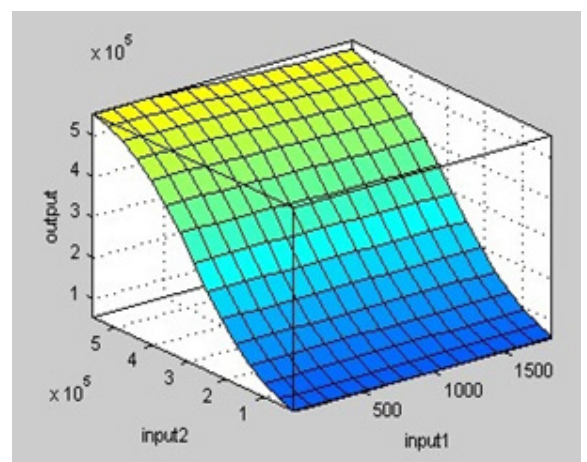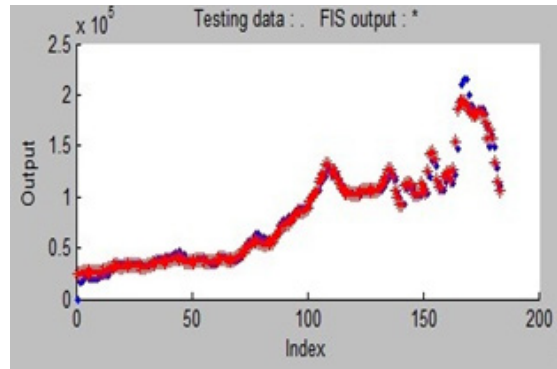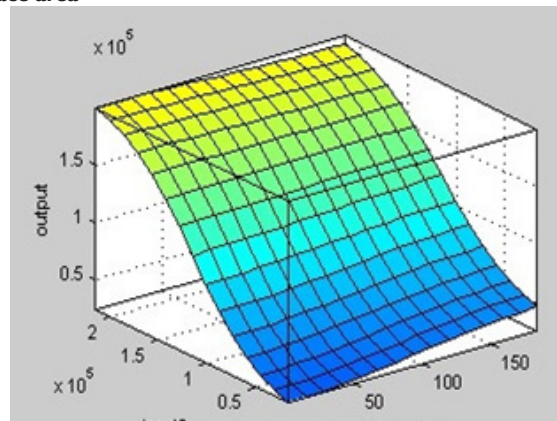
Figure 6a For the year 2012- Predicted values and surface area.

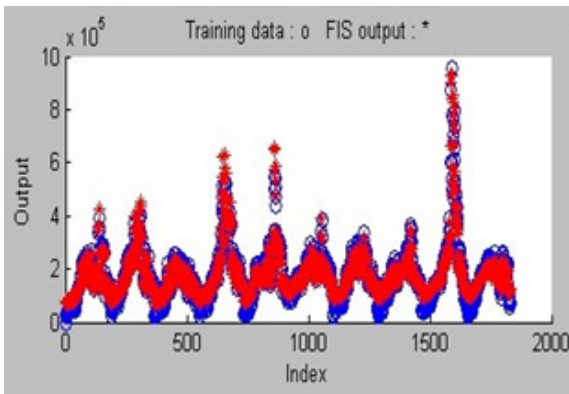**Predicted values**

Surface area

Figure 7a For the year 2012- Predicted values and surface area.
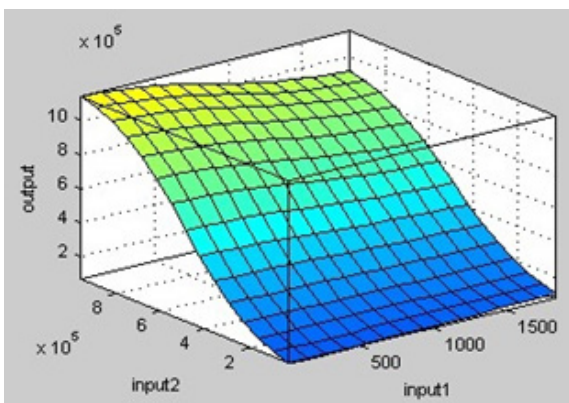
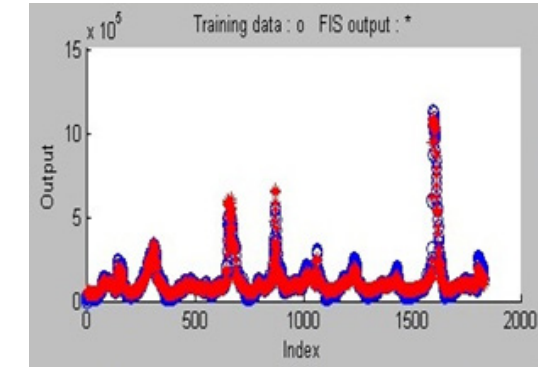**Predicted values**

Surface area

Figure 7b For the years 2002 to 2011- Predicted values and surface area.

Predicted values

Surface area

Figure 8a For the year 2012- Predicted values and surface area.
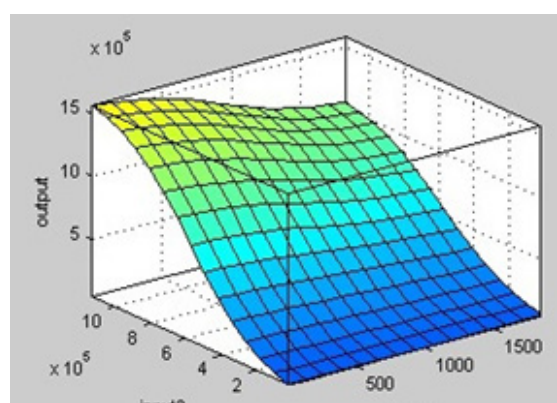
Predicted values

Surface area

Figure 8b For the years 2002 to 2011- Predicted values and surface area.

Outcomes and analysis from ARIMA

For tarbela dam results

Here, we have done calculations for Tarbela Dam below. Tables 5a and 5b are completely showing the results of the daily stream flow prediction for the whole data from the year 2002 to 2012 Figure 9.
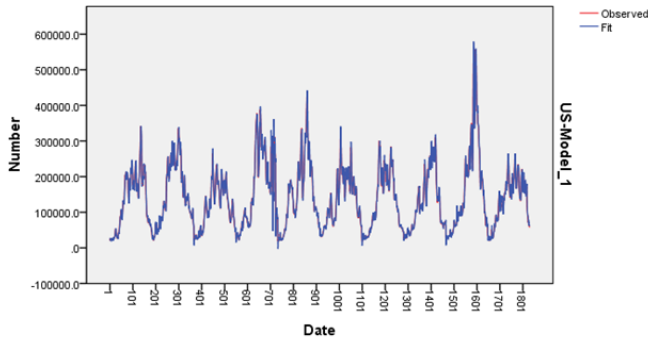

Figure 9 From the year 2002 to 2012-Predicted values of Stream flows for Tarbela Dam.

Model description

		Model type
Model ID	US Model_I	ARIMA (0,1,10)

For chashma barrage results

Now the calculations for Chashma Barrage are below. Table 6a and 6b are showing the outcomes of the daily stream flow prediction for the whole data from the year 2002 to 2012 Figure 10.
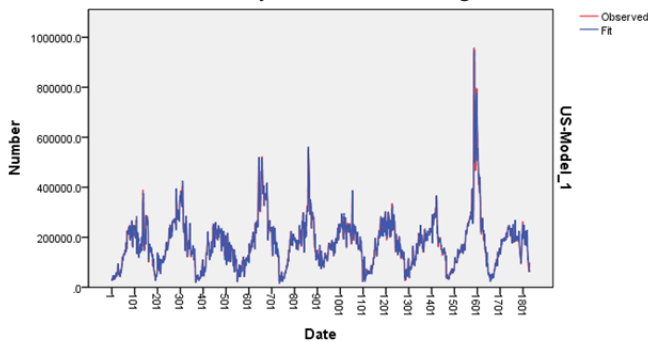

Figure 10 From the year 2002 to 2012-Predicted values of Stream flows for Chashma Barrage.

Model description

		Model type
Model ID	US Model_I	ARIMA (0,1,16)

For sukkur barrage results

Similarly, the calculations for Sukkur Barrage are below. Table 7a and 7b are showing the complete statistics of the daily stream flow prediction for the whole data from the year 2002 to 2012 Figure 11.

Model description

		Model type
Model ID	US Model_I	ARIMA (0,1,16)

Discussion on flood analysis results

Now the results of the different models are applied as in Table 2a to 7b. The results obtained by ANFIS modeling are much faster and better than the ARIMA model application.

First, take a look at the ANFIS model by comparing the three gauge stations results as we can see the calculating error RMSE is showing in the ANFIS and the error reduces by increasing the inputs than MFs is shown in this ANFIS model. It is very much surprising that with the increase in the membership functions the error RMSE is slowly decreasing while the increase in the number of inputs helps to make an efficient decrease in the calculated error. So we can easily interpret that increase in the inputs is a much better option for forecasting future values.
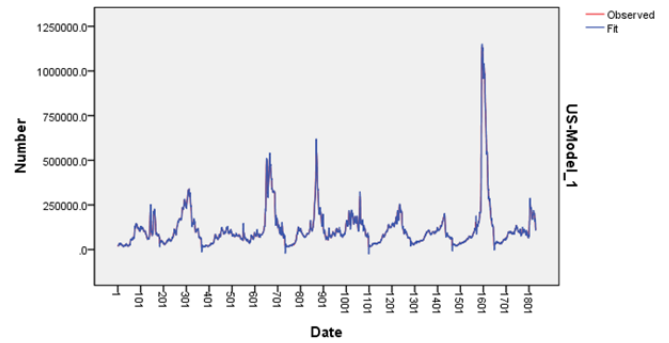

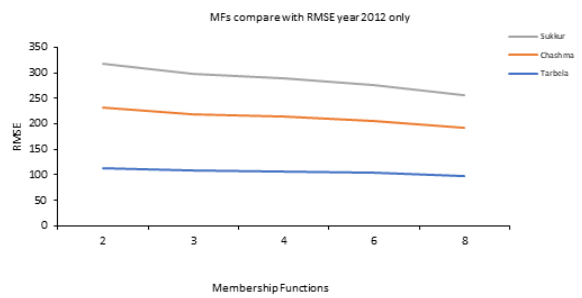**Figure 11** From the year 2002 to 2012-Predicted values of Stream flows for Sukkur Barrage.

Figure 12a Structure for 2-inputs d_t & d_{t-1}.
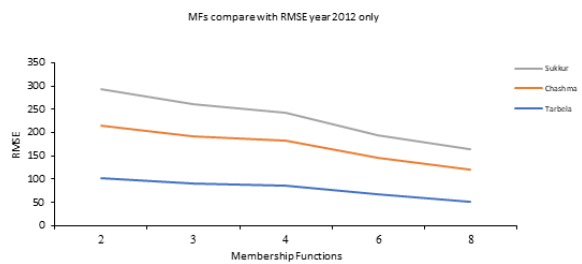

Figure 12b Structure for 3-inputs d_t, d_{t-1} & d_{t-2}.

Moreover, we have found that the outcomes of the three gauge stations Tarbela, Chashma, and Sukkur indicate the good results it can be attained by the structure of the ANFIS model. We observed the best outcomes from the year 2012 for testing data with the minimum errors for all stations. In Tarbela Dam, we used the data inputs d_t, d_{t-1}, d_{t-2}, and eight membership functions (MFs) as seen in Table 2a. Similarly, Table 3a is for the Chashma barrage and Table 4a is for the Sukkur barrage with the same data input structures where the flood risk is very high. And also in Table 4b can be seen an enormous decrease in the error i.e. 87.92 for 10 years' data using the same input structures with the same MFs as compared with the input structure of two d_t, d_{t-1} input, and two MFs as 170.087. Now we can elaborate on these

outputs as more inputs and MFs give the best results with minimum errors as 44.276 RMSE from the year 2012 by Sukkur barrage where always a high risk of the flood as shown in the Figure 12a and 12b for comparing the three stations results by the increase in the input data.

On the other hand, we used the ARIMA model to compare with ANFIS results and obtained a high amount of errors (RMSE) readings for all stations like Tarbela, Chashma and Sukkur barrages as 13513.881, 23598.722, and 11174.414 RMSEs respectively which is too large in amount.

Table 5a Model Fit

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile						
					5	10	25	50	75	90	95
Stationary R-squared	0.268	.	0.268	0.268	0.268	0.268	0.268	0.268	0.268	0.268	0.268
R-squared	0.977	.	0.977	0.977	0.977	0.977	0.977	0.977	0.977	0.977	0.977
RMSE	13513.88	.	13513.88	13513.88	13513.88	13513.88	13513.88	13513.88	13513.88	13513.88	13513.88
MAPE	6.932	.	6.932	6.932	6.932	6.932	6.932	6.932	6.932	6.932	6.932
MaxAPE	326.571	.	326.571	326.571	326.571	326.571	326.571	326.571	326.571	326.571	326.571
MAE	7990.652	.	7990.652	7990.652	7990.652	7990.652	7990.652	7990.652	7990.652	7990.652	7990.652
MaxAE	163891.6	.	163891.6	163891.6	163891.6	163891.6	163891.6	163891.6	163891.6	163891.6	163891.6
Normalized BIC	19.068	.	19.068	19.068	19.068	19.068	19.068	19.068	19.068	19.068	19.068

Table 5b Model Statistics

Model	Number of Predictors	Model Fit statistics				Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	RMSE	MAPE	Normalized BIC	Statistics	DF	Sig.	
US-Model_1	0	0.268	13513.88	6.932	19.068	17.372	8	0.026	0

Table 6a Model Fit

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile						
					5	10	25	50	75	90	95
Stationary R-squared	0.122	.	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
R-squared	0.946	.	0.946	0.946	0.946	0.946	0.946	0.946	0.946	0.946	0.946
RMSE	23598.72	.	23598.72	23598.72	23598.72	23598.72	23598.72	23598.72	23598.72	23598.72	23598.72
MAPE	10.604	.	10.604	10.604	10.604	10.604	10.604	10.604	10.604	10.604	10.604
MaxAPE	331.573	.	331.573	331.573	331.573	331.573	331.573	331.573	331.573	331.573	331.573
MAE	15563.31	.	15563.31	15563.31	15563.31	15563.31	15563.31	15563.31	15563.31	15563.31	15563.31
MaxAE	231219.3	.	231219.3	231219.3	231219.3	231219.3	231219.3	231219.3	231219.3	231219.3	231219.3
Normalized BIC	20.208	.	20.208	20.208	20.208	20.208	20.208	20.208	20.208	20.208	20.208

Table 6b Model Statistics

Model	Number of Predictors	Model Fit statistics				Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	RMSE	MAPE	Normalized BIC	Statistics	DF	Sig.	
US-Model_1	0	0.122	23598.72	10.604	20.208	1.057	2	0.59	0

Table 7a Model Fit

Fit Statistic	Mean	Minimum	Maximum	Percentile						
				5	10	25	50	75	90	95
Stationary R-squared	0.506	0.506	0.506	0.506	0.506	0.506	0.506	0.506	0.506	0.506
R-squared	0.992	0.992	0.992	0.992	0.992	0.992	0.992	0.992	0.992	0.992
RMSE	11174.41	11174.41	11174.41	11174.41	11174.41	11174.41	11174.41	11174.41	11174.41	11174.41
MAPE	5.083	5.083	5.083	5.083	5.083	5.083	5.083	5.083	5.083	5.083
MaxAPE	338.483	338.483	338.483	338.483	338.483	338.483	338.483	338.483	338.483	338.483
MAE	4871.694	4871.694	4871.694	4871.694	4871.694	4871.694	4871.694	4871.694	4871.694	4871.694
MaxAE	246516.8	246516.8	246516.8	246516.8	246516.8	246516.8	246516.8	246516.8	246516.8	246516.8
Normalized BIC	18.663	18.663	18.663	18.663	18.663	18.663	18.663	18.663	18.663	18.663

Table 7b Model Statistics

Model	Number of Predictors	Model Fit statistics				Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	RMSE	MAPE	Normalized BIC	Statistics	DF	Sig.	
US-Model_1	0	0.506	11174.41	5.083	18.663	34.701	13	0.001	0

Conclusion

We have executed an Adaptive Neuro-Fuzzy Inference System (ANFIS) at all three gauge stations to expect the cyclic behavior of river flow discharges. Different Input variables were applied with different membership functions by using two types of neuro-fuzzy systems operated 5-times with 2, 3, 4, 6, and 8 MFs and with 2 & 3 data inputs. For this purpose, we accumulated ten years' stream flow discharge data for these three gauge stations along River Indus flow and used it as training data. Another one is executed for a one-year stream flow as testing data. The system was executed for different levels. We obtained better results by increasing the no. of inputs instead of increasing the no. of membership functions to the fuzzy network. By comparing both the models ANFIS & ARIMA, we can conclude our outcomes based on RMSE values obtained from the different models and by graphs that the model ANFIS is the better option to predict and forecast floods by the recorded daily streamflow time series data as ANFIS gave us the minimum value of RMSE mentioned in the above tables. There is a comparison between the observed and the predicted data values. We can say that the model of ANFIS can be utilized in the future as this is very adaptable, fruitful, and has many possibilities of integrating the real world's nature for the time series analysis.

Acknowledgments

Conflicts of interest

The author declares there is no conflict of interest.

References

1. GEP Box, GM Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, Oakland, CA. 1970.

2. RD Valencia, JC Schaake. Disaggregation processes in stochastic hydrology. *Water Resour Res*. 1973;9(3):58–585.

3. EH Mamdani, S Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *Int J Man Mach Stud*. 1975;7(1):1–13.

4. Y Tsukamoto. *An Approach to Fuzzy Reasoning Method*. In: Gupta M, et al., editors. Advances in Fuzzy Set Theory, North-Holland, Amsterdam. 1979.

5. JD Salas, RA Smith. Physical basics of stochastic models of annual flows. *Water Resour Res*. 1981;41(4):428–430.

6. JR Stedinger, MR Taylor. Synthetic stream-flow generation model verification and validation. *Water Resource Res*. 1982;l(18):909–918.

7. D Burn, E McBean. River flow forecasting model for the Sturgeon River. *J Hydraulic Eng*. 1985;111(2):316–333.

8. T Takagi, M Sugeno. Fuzzy identification of systems and their application to modeling and control. *IEEE Trans Syst Man Cybernet*. 1985;15(1):116–132.

9. DE Rumelhart, GE Hinton, RJ Williams. Learning representations by back-propagating errors. *Nature*. 1986;323:533–536.

10. M Sugeno, GT Kang. Structure identification of the fuzzy model. *Fuzzy Sets Syst*. 1988;28:15–33.

11. JD Kalma, GP Laughlin. Flood risk mapping for landscape planning: a methodology. *International Journal of Theoretical and Applied Climatology*. 1990;42:41–51

12. JSR Jang. Rule extraction using generalized neural networks. In Proceedings of the fourth IFSA World Congress. *Volume for Artificial Intelligence*. 1991;4:82–86.

13. JSR Jang. ANFIS: Adaptive network-based fuzzy inference systems. *IEEE Trans. Syst., Man Cybernet*. 1993;23(3):665–685.

14. Jang JSR, Sun CT, Mizutani E. *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Prentice-Hall, Upper Saddle River, NJ. 1997.

15. M Sugeno, T Yasukawa. A fuzzy-logic-based approach to qualitative modeling. *IEEE Transactions on Fuzzy Systems*. 1993;1:7–31.

16. H Awwad, J Valdes, P Restrepo. Streamflow forecasting for Han River Basin, Korea. *J Water Resour Planning Manage*. 1994;120(5):651–673.

17. M Brown, C Harris. *Neuro-fuzzy Adaptive Modeling and Control*. Prentice Hall. 1994.

18. M El-Fandy, Z Ashour, S Taiel. Time series models are adaptable for forecasting Nile floods and Ethiopian rainfalls. *Bull Am Meteorol Soc*. 1994;75(1):1–12.

19. N Karunanithi, WJ Grenney, D Whitley. Neural networks for river flow prediction. *J Comput Civ Eng ASCE*. 1994;8(2):210–220.

20. JSR Jang, CT Sun. Neuro-fuzzy modeling and control. *Proceedings IEEE*. 1995;83(3):378–406.

21. T Ojala. *Neuro-Fuzzy systems in control*. M Sc. Thesis, Tempere University of Technology, Tampere, Finland. 1995.

22. K Smith. *Environmental hazards*. Rutledge, London. 1996.

23. Keskin ME, Taylan D, Terzi O. Adaptive neural-based fuzzy inference system (ANFIS) approach for modeling hydrological time series. *Hydrological Sciences Journal*. 2006;51(4):588–598.

24. KP Sudheer, PC Nayak, DM Rangan. *Rainfall-runoff modeling using artificial neural network technique, Report No*. CS/AR-16/1999-2000, National Institute of Hydrology, Roorkee, India. 2000.

25. LH Xiong, AY Shamseldin, KM O'Connor. A nonlinear combination of the forecasts of rainfall-runoff models by the first order Takagi-Sugeno fuzzy system. *Journal of Hydrology*. 2001;245(1–4):196–217.

26. F Nawaz, M Shafique. *Data integration for flood risk analysis by using GIS/RS as tools*. In proceedings of MAP Asia, Second Annual Asian Conference and Exhibition in the field of GIS, GPS, Aerial Photography, and Remote Sensing. 2003.

27. PC Nayak, KP Sudheer, DM Rangan, et al. A neuro-fuzzy computing technique for modeling hydrological time series. *J Hydrol*. 2004;291:52–66.

28. TE Ologunorisa, MJ Abawua. Flood risk assessment: A review. *App Sci Environ Mgt J*. 2005;9(1):57–63.

29. G Khan. *Flood hazard assessment and mitigation along River Indus from Chashma Barrage to Sukkur Barrage using satellite image*. M.Phil. thesis, Institute of Space and Planetary Astrophysics, The University of Karachi. 2007.

30. B Khan, MJ Iqbal, MAK Yosufzai. Flood risk assessment of River Indus of Pakistan. *Arab Journal of Geoscience*. 2009;10: 1007/s12517-009-0110-9.

31. *Annual Flood report, 2010, from Federal Flood Commission, Department of Ministry of Water and Power, and Government of Pakistan-Islamabad*. 2010.

32. Daniyal Hassan, Rakhshinda Bano, Kamran Ansari, et al. *Hydrology of Upper Indus Basin*. 2010.

33. Cheng CT, Lin JY, Sun YG, et al. *Long-term prediction of discharges in Manwan hydropower using adaptive-network-based fuzzy inference systems models, Advances in Natural Computation, Pt 3, Proceedings*. Lecture Notes in Computer Science. Springer-Verlag Berlin, Berlin, 2005. p. 1152–1161.

34. He S, Lunnen JC, Zia N, et al. The pattern of presenting complaints recorded as near-drowning events in emergency departments: a national surveillance study from Pakistan. *BMC emergency medicine*. 2015;15(2):1–9.

35. Ateeq-Ur-Rehman S, Bui MD, Rutschmann P. Detection and estimation of sediment transport trends in the upper Indus River during the last 50 years. *Hydropower: A Vital Source of Sustainable Energy for Pakistan*. 2017;1–6.

36. Tayyab M, Ahmad I, Sun N, et al. Application of integrated artificial neural networks based on decomposition methods to predict streamflow at upper Indus basin, Pakistan. *Atmosphere*. 2018;9(12):494.

37. Nazir HM, Hussain I, Faisal M, et al. Improving the prediction accuracy of river inflow using two data pre-processing techniques coupled with a data-driven model. *Peer J*. 2019;7:e8043.