

ARTIFICIAL NEURAL NETWORK MODELING FOR ANNUAL PEAK FLOWS: A CASE STUDY

AHMAD, I.^{1,2*} – ALMANJAHIE, IBRAHIM M.¹ – HAMEEDULLAH² – CHIKR-ELMEZOUAR, Z.¹ – LAKSACI, A.¹

¹*Department of Mathematics, College of Science, King Khalid University, 61413 Abha, Kingdom of Saudi Arabia*

²*Department of Mathematics and Statistics, Faculty of Basic and Applied Sciences, International Islamic University, 44000 Islamabad, Pakistan*

**Corresponding author
e-mail: Ishfaq.ahmad@iiu.edu.pk*

(Received 15th Feb 2019; accepted 8th Apr 2019)

Abstract. The data of annual peak flows in river system is not straight forward but rather complex function of hydrology and geography of the area. In such composite situations, models based on soft computing approaches such Artificial Neural Network (ANN) may increase understanding of the hidden hydrological processes. The primary objective of this study is to determine the best approach among Traditional Artificial Neural Network (T-ANN) and Wavelet Artificial Neural Network (W-ANN) using Annual Peak flows data of Mangla site on River Jhelum, Pakistan. The results reveal the better performance of W-ANN approach under db2 scheme. Further, to justify the authenticity of W-ANN we compared it with the frequency analysis by fitting best probability distribution and non-parametric kernel density estimation (KDE) using Gaussian kernel approach for the same data. Results reveal that probability distribution modeling approach by fitting the Generalized Logistic (GLO) distribution being the most suitable model provides comparable results with W-ANN-db2 approach but outperforms than non-parametric KDE. Overall, the performance of W-ANN-db2 is better in the present study. The findings of the study are useful for better policy implications in water resources management.

Keywords: *flood frequency analysis, water resources management, probability models, kernel density estimation, non-parametric*

Introduction

We are living in the age where we have more awareness than ever about the environmental events around us, including the brittleness of ecosystems, the incompatible needs of human culture, and the complications of geologic records. Scientists are looking for, to create, upgrade, and revise the theories and mathematical models of the Earth's dynamics at an unprecedented rate. Rather than traditional modeling techniques, the Artificial Neural Network (ANN) methodology has the prospective aspects to see the problems in a simpler and more modular form. For sustainable development of the societies it is an urgent need to make predictions with more accuracy of different environmental events including floods, earthquakes, droughts, air pollution, wind gusts and others. Since last few years ANN are widely applied in different fields such as Image Processing, Signal Processing, power systems, medical studies, pattern recognition and hydrology among others (for example, see Rabuñal, 2005; Aziz et al., 2016; Jimeno-Sáez et al., 2017; Kumar et al., 2015; Rahman et al., 2018; Requena et al., 2018 among others). ANN modeling has the capability of learning from the data. They can be proved as time saving for development of the models and especially useful for nonlinear and complicated processes which cannot be modeled through traditional methods. Streamflow forecasting is a challenging task for

water resources engineers and managers. Reliable forecasts of streamflows make an efficient operation of water resources systems within legal, political, technical and economical priorities. A forecasting system that considers all important spatial and temporal variation of the whole streamflow regime, delivers a good basis for appropriate control and management of the water resources system.

The principle parameters for flood forecasts are scale and frequency of events. Since the annual peak flows can be used to give flood forecasts. Statistical studies in view of probabilistic methods can be embraced to tell about the probability and frequency of peak flows that cause flooding. The principal parameters for flood forecasts in probabilistic approach are frequency and magnitude of events (for example see Ahmad et al., 2015; Rahman et al., 2013; Drissia et al., 2019; Cassalho et al., 2019; Mostofi et al., 2019). Mostly, the Hydrologic time series are autocorrelated. However annual precipitation and peak flows are generally uncorrelated (Salas et al., 1993). Rarely, substantial autocorrelation might be the result of trends as well as movements in the time series (Salas et al., 1988). There is lot of current literature which discusses about aspects related to flood occasions (Agarwal et al. 2005; Rahimi and Abbaspour, 2011; Guresen et al., 2011.) Peak flows in the stream or waterway is a composite capacity of hydrology of the area and land designs in a given stream basin or waterway being considered as are hard to evaluate precisely. Generally the models which use streamflow data along with many other hydrologic and climatic factors for example precipitation, temperature, and snow water in a completely statistical or stochastic manners are black-box type models (see for example Hipel and MeLeod, 1994).

In the current study, we have applied ANN and W-ANN models to only one variable that is peak flows of Mangla site and compared the results of W-ANN (being the better one) with probability distribution modeling approach and Non-Parametric KDE approach. ANN modeling for peak flows has been performed in some studies (see for example Harinarayan et al., 2015).

Materials and methods

Study area and data collection

The annual peak flows data of Mangla site situated at River Jhelam, Pakistan is being retrieved from Water and Power Development Authority (WAPDA) and Federal Flood Commission of Pakistan. The data used in this study (annual peak flows) have been recorded during monsoon season in Pakistan (from July to September). Monsoon is the peak season of rainfalls and peak flows in Pakistan. The river system of Pakistan begins from the snow-covered Himalayan and the Karakoram series. There are mainly five rivers mostly located in the province of Punjab. These rivers are Sutlej, Ravi, Jhelum, Indus and Chenab. River Jhelum originates from south eastern part of Kashmir and its width is approximately 774 km and it meets with Chenab River at Trimmu. Jhelum also passes through Srinagar before entering Pakistan. Number of barrages and dams have been constructed on River Jhelum. One of them is Mangla dam constructed in 1967. It has a stowage dimensions of approximately 5.9 million acre-feet.

Artificial neural network (ANN)

ANN approach is based on function of the human brain and nervous system. ANNs trace the associations between two types of factors, i.e. inputs and outputs of a task by

interlinks of nodes that activate and operate through connecting weights using different functions called “training samples”, and in the result we can determine the procedures and relationship between the inputs and outputs. The processing of every neuron is performed in dual steps, in first step, the weighted summation of the inputs is brought and in second they are trailed by the importance of activation function (Bishop, 1995).

Artificial Neural Network are commonly applied by using electrical mechanisms or by imitation on a computer system. They are categorized by the designs of networks amongst the nodes (named as architecture), the procedures of defining weights on the networks (called their learning procedure), the activation function, and the number of layers (single layer, bilayer, and multilayer). When overall signals of a network move in only one direction, the ANN is known as Feed-Forward Network (FFN), otherwise they are known as Recurrent Networks. Generally, the artificial neural networks used in hydrological models are Feed-Forward Multilayer Perceptron’s (FFMLPs), which include three layers, and employ monotonic activation functions in all layers. Their weights are generally found by organized training using back-propagation process. The performance and learning capability of artificial neural network depend upon the fitness of its architecture which is needed to be pre-specified more precisely, the number and configuration of its hidden nodes see for example (Hertz, 2018; Denis, 2009).

Composite non-linear problems such as time series data prediction can be solved through ANN being non-linear mathematical model. Multilayer Perceptron (MLP) is the type of Artificial Neural Network which is used to adjust the connection weights. MLP comprises three layers: input layer; hidden layer; and output layer. MLP has only one input and output layer but more than one hidden layer. Each of these layers comprises nodes. These nodes are associated to each other. Input layer can contain that number of nodes while output layer, must have only one node to get output. It is up to user selection how to use hidden layer. There must be an additional dummy neuron as bias node in input layer. Bias node acts as one of the players in the network. The activation function is used to calculate the arriving value and gives the value in output layer for the actual computational process occurrence. Input layer applies linear transfer function. However, a sigmoid function is normally used in hidden and output layer. Following the activation function, we can determine the connection weight by data training. The procedure for the adjustment of weight begins in the output layer and carries backward to the input layer. Until the performance of target is obtained the process of feed-forward and back propagation algorithm continues. The weights of the values are determined when the data training process is completed. Following this, a single feed forward computation is employed. Results are then assessed by conducting performance measures between actual and predicted values (for detail see for example Bishop, 1995; Singh et al., 2004).

This study used Levenberg Marquardt (LM) procedure to streamline the biases and weights in the system. LM procedure is more effective and quicker than the traditional gradient descent procedure (Huang et al., 2015; Kisi and Shiri, 2011). *Figure 1* shows the Multilayer Perceptron with three layers as stated before.

The *Figure 1* shows a three-layer feed forward ANN. The first layer, called the input layer, comprises I nodes and connected with the input variables. This layer performs no computation but is used to distribute I inputs into the network. The last layer is connected to the output variables, and called the output layer. This layer consists of K output nodes. One or more layers of processing units that have no direct relation to the outside variable and are located between the input and output layers are called hidden

layers. An MLP may have many hidden layers. However, only one hidden layer is considered in *Figure 1*. This layer consists of J hidden nodes. In each layer, the processing elements are called nodes or neurons. Each of these neurons connected to all the neurons of adjoining layers. The parameters linked with these connections are known as weights. Information used by the network to solve a problem is represented by weights. Generally, all connections are ‘feed forward’ that allow information to transfer from an earlier layer to the next consecutive layer. The neurons in a layer are not interconnected, and neurons in non-adjointing layers are also not connected, however, for recurrent neural networks this is not the case (Zurada, 1992).

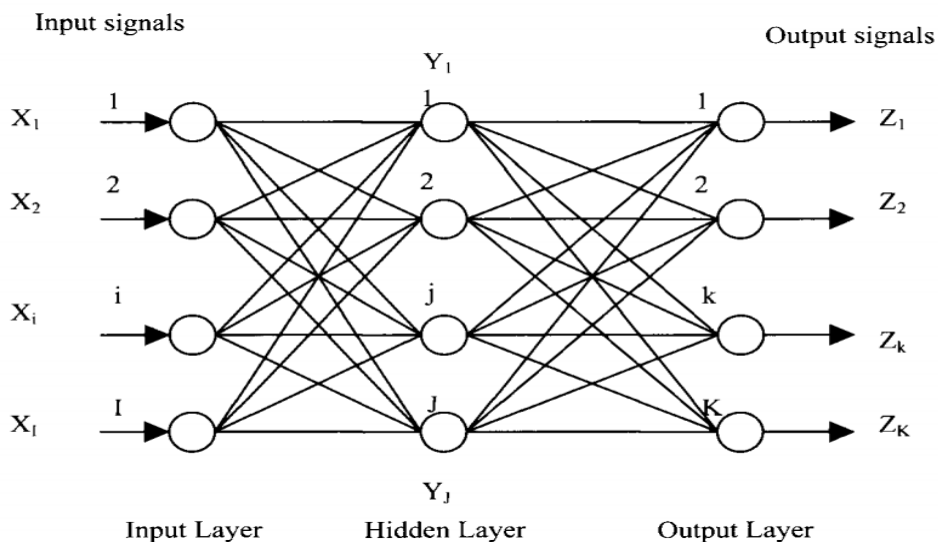


Figure 1. Multilayer perceptron with three layers artificial neural network model

The architecture of a typical neuron is shown in the *Figure 2*. Every hidden neuron j gets total I arriving signals (X_i , i.e. $X_1, X_2, \dots, X_i, \dots, X_I$) from any neuron i in the preceding layer (for instance, from Input layer to hidden layer). Connected with every arriving signal (X_i), a weight (W_{ij}) is connected between layer i and j . The effective arriving signal (net_j) to neuron j is the sum of weights of all the arriving signals as follows in *Equation 1*:

$$net_j = \sum_{i=1}^I W_{ij} X_i \quad (\text{Eq.1})$$

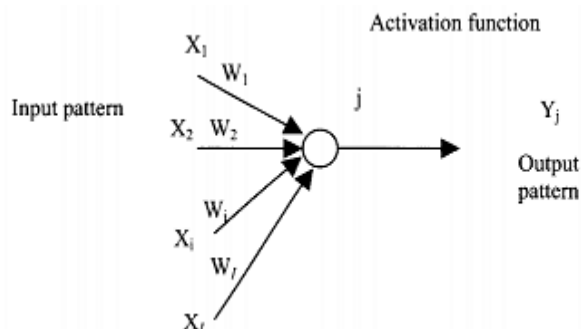


Figure 2. The architecture of a typical neuron

This effective arriving signal (net_j) passes over a non-linear activation function (also called a transfer function or threshold function) that produces the outgoing signal which is called activation or activity level (Y_j) of the neuron. Usually, a neuron sends its activation as a signal to several other neurons. It is pertinent to note that at one time only one signal can be sent by a neuron, though that signal is programmed to numerous other neurons simultaneously.

Delta learning rule for feed-forward multilayer perceptron

The delta learning rule is also called the LMS (Least Mean Square) error back propagation training algorithm. In the feed forward propagation training Input patterns are entered sequentially. If a pattern is entered and its classification or association comes out to be inaccurate, the value of LMS error is minimized through the adjustment of the synaptic weights as well as the thresholds. Therefore, the purpose of this algorithm is to adjust the synaptic weights w_{ij} ($i = 1, \dots, I$) and w_{jk} ($j = 1, \dots, J$), which connect input to hidden layer, and then hidden to output layer respectively. In the result, it will assure the minimization of the error function $\{E_k\}$ given in Equation 2. The instantaneous estimate of the mean-square error calculated at the output layer is defined as follows (Derks, 1995; Mukerji, 2009):

$$E_k = \frac{1}{2} (z_k - t_k)^2 \quad (\text{Eq.2})$$

where z_k the output from output layer and t_k is the target value. The outgoing signal (z_k) is a function of the activation as follows (Eq. 3):

$$z_k = f(net_k) \quad (\text{Eq.3})$$

The effective incoming signal net_k , comprising weighted sum signals from the hidden layer (y_j) and calculated as follows (Eq. 4):

$$net_k = \sum_{j=1}^J w_{jk} y_j \quad (\text{Eq.4})$$

The outgoing signal (y_j) is the result of the incoming signal net_j , being passed through the activation function, which is not necessarily the same as that for z_k , is given as follows (Eq. 5):

$$y_j = f(net_j) \quad (\text{Eq.5})$$

Again, the effective incoming signal net_j , comprising weighted sum signals from input layer (x_i) is calculated as Equation 6:

$$net_j = \sum_{i=1}^I w_{ij} x_i \quad (\text{Eq.6})$$

The derivative of the error functions with respect to w_{jk} , can be expressed after applying the chain rule, as follows in Equation 7:

$$\frac{\partial E_k}{\partial w_{jk}} = \frac{\partial E_k}{\partial Z_k} \frac{\partial Z_k}{\partial net_k} \frac{\partial net_k}{\partial w_{jk}} \quad (\text{Eq.7})$$

Finally, the updated weights are given in *Equation 8*:

$$w_{ij} = w_{ij} + \Delta w_{ij} \quad (\text{Eq.8})$$

Activation function

Artificial neural network is a procedure to sum the product of the connected weights, the input signals, and produce an output or activation transfer function. The activation function is the specific function for the input units. The nodes of a specific layer get the equivalent type of activation function. Generally, in every case a non-linear activation transfer function is used. There are several activation transfer functions employed in ANNs, but in this study we will focus on unipolar binary function or sigmoid function (S).

The unipolar binary function or sigmoid function (S)

The sigmoid function, as implemented by Hall and Minns (1996) and Dawson and Wilby (2001) is defined as in *Equation 9*:

$$S(\text{net}_j) = \frac{1}{1 + \exp(-\lambda \text{net}_j)} \quad (\text{Eq.9})$$

where $\lambda > 0$ in above sigmoid function. λ is in proportion to the nodes defining the steepness of the function, which in general is set equal to one. The term λnet_j , will lies between $(-\infty$ and $+\infty)$, whereas $S(\text{net}_j)$ will lies between (0 and 1).

Wavelet analysis

Two primary types of wavelet transforms are commonly used i.e. continuous wavelet transform (CWT) and Discrete Wavelet Transform (DWT). The CWT works with functions defined over the whole real axis while the DWT deals with functions defined over a range of integers. Wavelets are mathematical processes which represent a time-scale of the time series and their relationships to analyze time series when the data is non-stationary. Wavelet analysis is helpful in the long-time intervals where the frequency information's are low and also in the short time intervals where the frequency information is high.

Wavelet is a small wave function denoted by the notation $\psi(\cdot)$. The range of a wavelet mother function ($\psi(\cdot)$) lies between $-\infty$ and ∞ , and it must fulfill the following given properties:

The value of the integral of $\psi(\cdot)$ is always equal to zero as shown in *Equation 10*:

$$\int_{-\infty}^{+\infty} \Psi(u) du = 0 \quad (\text{Eq.10})$$

The integral of the square of $\psi(\cdot)$ is unity as in *Equation 11*:

$$\int_{-\infty}^{+\infty} \Psi^2(u) du = 1 \quad (\text{Eq.11})$$

Admissibility Condition of $\Psi(f)$ is given in *Equation 12*:

$$C_{\Psi}(t) = \int_0^{\infty} \Psi^* \frac{[\Psi(f)]^2}{f} df \quad \text{satisfies} \quad 0 < C_{\Psi} < \infty \quad (\text{Eq.12})$$

Haar created a wavelet function in 1910 (*Fig. 3*), named Haar wavelet function (Ogden, 2012) and considered one of the oldest wavelet functions, given in *Equation 13*:

$$\Psi^{(H)}(u) = \begin{cases} +1 & \text{if } 0 \leq u < \frac{1}{2} \\ -1 & \text{if } \frac{1}{2} \leq u < 1 \\ 0 & \text{else} \end{cases} \quad (\text{Eq.13})$$

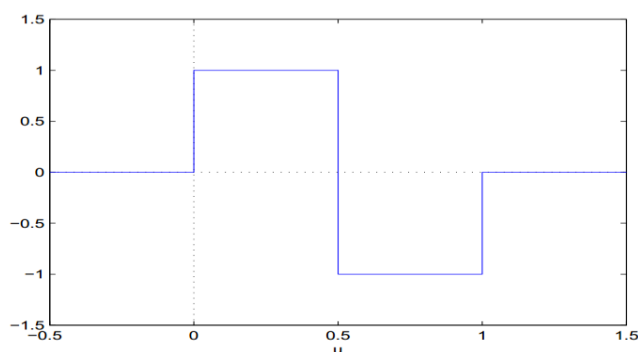


Figure 3. The Haar wavelet function

It is a discrete wavelet and looks like a step function. Harr wavelet and Daubechies (db) are the same. Daubechies family wavelets are written as dbN, where N stands for the order of Daubechies family. There is more flexibility in wavelet analysis due to which we can choose a specific Wavelet for matching the type of function under analysis.

The function $\psi(\cdot)$ is the mother wavelet and a double wavelet index family can be constructed by *Equation 14*:

$$\Psi_{\lambda,t}(u) = \frac{1}{\sqrt{\lambda}} \Psi\left(\frac{u-t}{\lambda}\right) \quad (\text{Eq.14})$$

where $\lambda > 0$ and t is finite.

The standardization on the right-hand side of *Equation 14* is:

$$\|\Psi_{\lambda,t}\| = \|\Psi\| \quad \text{for all } \lambda, t.$$

Functions are considered linear combination where the information is secure and make no losses to those functions which are in discrete form or as an integral in the continuous form of the selected wavelet family.

Wavelet-ANN conjunction model for flood events prediction

The wavelet analysis is connected to the Artificial Neural Network approach for annual peak flows prediction in this research, also depicted in *Figure 4*. First of all, the

observe annual peak flows time series, Q were decayed into several multi-frequencies time series including details (high frequency, low scale), such as, $Q_{D1}(t)$; $Q_{D2}(t)$; $Q_{Di}(t)$, and approximate (high scale, low frequency) such as, $Q_{Da}(t)$ by DWT. The symbol D_i shows the Wavelet-ANN Model for Flood prediction, level i decayed details time series and a denote approximation time series.

However, the decayed $Q(t)$ time series were used as inputs to ANN model. The original data of time series in the next step, $Q(t + 1)$ is output of the Wavelet-ANN conjunctural model. The graphical representation of above methodology is shown in *Figure 4*.

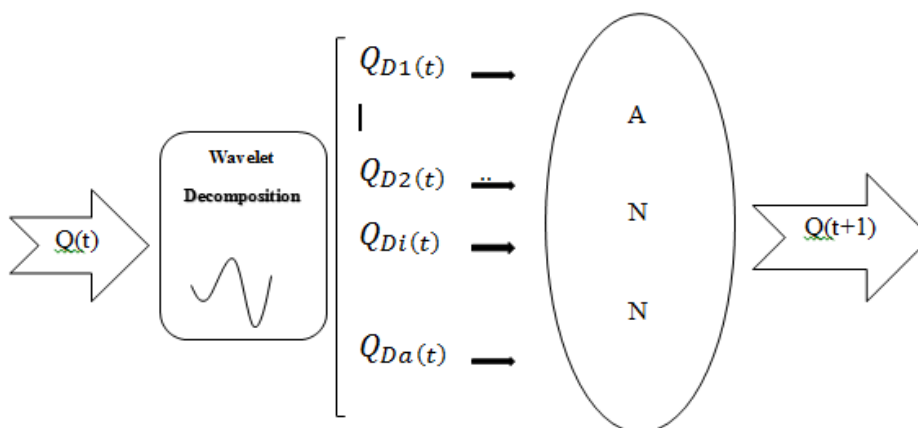


Figure 4. Schematic diagram of wavelet-ANN conjunction model

Criteria for the evaluation of a model

To evaluate the consistency of the models, it is compulsory to check the capabilities of ANN in both training and testing period (see for example Beven, 2011). Several statistical techniques are generally used for this purpose. Some of the evaluations measures are given below, which will be used in this study.

Pearson correlation coefficient (R)

Correlation coefficient (R) designates the direction of the linear association among two variables (i.e., model input and output values). The R coefficient is achieved by dividing the co-variance of the two random variables by the multiplication of their standard deviations. The R is given by the formula given in *Equation 15*:

$$R = \frac{\sum_{i=1}^n (Y_{mi} - \hat{Y}_{mi})(Y_{ci} - \hat{Y}_{ci})}{\sqrt{\sum_{i=1}^n (Y_{mi} - \hat{Y}_{mi})^2 \sum_{i=1}^n (Y_{ci} - \hat{Y}_{ci})^2}} \quad (\text{Eq.15})$$

where Y_{mi} indicate the observed values and Y_{ci} indicate the model or predicted values of annual peak flows at particular site at time i ., \hat{Y}_{mi} and \hat{Y}_{ci} are the average values of Y_{mi} and Y_{ci} respectively.

If the association is +1 then there is a perfect positive linear relationship between actual and estimated values while a value of -1 refers to perfect negative linear relationship between them. If R comes with 0 value, then it indicates no linear relationship between them.

Root mean square error (RMSE)

The positive square root of the mean/average of the square of all the errors. RMSE measures the spread/ variation of these residuals. RMSE is a commonly used measure to see the variances between the predicted values and estimated values with a range from 0 to ∞ . When $Y_{mi} = Y_{ci}$, then $RMSE = 0$, which implies that the model is perfectly fitted. The smaller RMSE shows that the model is best fitted. The formula of RMSE is given by Equation 16:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_{mi} - Y_{ci})^2} \quad (\text{Eq.16})$$

Mean absolute error (MAE)

MAE calculates the average size of the errors, without caring about the direction of the error. It is average of absolute differences between predicted and estimated values. The minimum value of MAE indicates that the model is best fitted. The formula of MAE is shown as given in Equation 17:

$$MAE = \frac{1}{n} \sum_{j=1}^n |Y_{mi} - Y_{ci}| \quad (\text{Eq.17})$$

Nash-Sutcliffe coefficient

To evaluate the prediction of a hydrological model this co-efficient is frequently used. It is proposed by Nash-Sutcliffe coefficient in 1970. The Nash-Sutcliffe coefficient E is defined by Equation 18:

$$E = 1 - \frac{\sum_{i=1}^n (Y_{mi} - Y_{ci})^2}{\sum_{i=1}^n (Y_{mi} - \bar{Y}_{mi})^2} \quad (\text{Eq.18})$$

The range of E lies between $-\infty$ and 1. A value of $E = 1$ indicates the perfect match among observed and predicted values while a value of 0 implies that the predictions of model are as correct as the average of the actual data. However the value of $-\infty < E < 0$ arises in case of week prediction. For more details see Grossmann (2010).

Index of agreement (IOA)

The term Index of Agreement established by Willmott (2012) as a standard quantity to evaluate the prediction error of a model and it lies between 0 and 1. The value of $IOA = 1$ shows an accurate prediction while of 0 shows no agreement at all. It is defined as in Equation 19:

$$IOA = 1 - \frac{\sum_{i=1}^n (Y_{mi} - Y_{ci})^2}{\sum_{i=1}^n (|Y_{mi} - \bar{Y}_{mi}| + |Y_{ci} - \bar{Y}_{ci}|)^2} \quad (\text{Eq.19})$$

Non-parametric kernel density estimation

We also consider non-parametric kernel density estimation (KDE) approach for comparison with ANN approach. In Statistics, KDE is a well-known nonparametric method, which is used to estimate the probability density estimation (PDF) of a random variable. KDE is a principal data smoothing issue where decisions about the population are made, on the basis of finite data sample. Cross validation (CV) smoothing technique

being more efficient is used in this study to estimate the smoothing parameter of the nonparametric KDE for annual peak stream flows of Mangla site. By using the estimates of smoothing parameters we estimate the nonparametric kernel density using Gaussian kernel (for details see Shabri, 2002; Lall, 1995; Silverman, 2018; Gramacki, 2018). Let Y be a random variable with PDF. The PDF f provides the actual behavior of the distribution of Y . The probability density function for a certain range is as follows in Equation 20:

$$\int_a^b f(y)dy \quad \forall a < b \quad (\text{Eq.20})$$

For parametric approach, it is always necessary to check the all fundamental assumptions but in non-parametric KDE, we move towards the assumption free estimation methodology.

Let K denotes the kernel function which holds the following condition given in Equation 21:

$$\int_{-\infty}^{+\infty} K(y) = 1 \quad (\text{Eq.21})$$

The kernel estimator can be defined as in Equation 22:

$$\hat{f}(y) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{y - Y_i}{h}\right) \quad (\text{Eq.22})$$

where h denotes the smoothing parameter or bandwidth or window width. The kernel estimator is the sum of bumps placed at the sample values. Further, kernel function K and bandwidth h determine the shape of the bumps and width for the kernel estimator, respectively. In NP kernel density estimation, calculation of smoothing parameter h is crucial. If the value of h is too small then used kernel estimator leads to larger variance and smaller bias. The value of smoothing parameter or bandwidth h has to be selected optimally. We have opted Cross Validation (CV) method in this study.

At site flood frequency analysis by fitting appropriate probability model

For fitting probability model to peak flows (*at-site flood frequency analysis*), we considered method of L-moments for estimation of parameters of different probability models and also for different goodness of fit tests. Method of L-moments is considered as more robust in the presence of extreme observations. After fulfilment of the basic assumptions of randomness, stationarity, independence and homogeneity through different statistical tests, the next step is to choose the best fit probability distribution among the most commonly used probability models in hydrology such as generalized Logistic (GLO), Generalized Extreme Value (GEV), Generalized Pareto (GPA), Pearson Type 3 (PE3), an Generalize Normal (GNO) (for example see Ahmad et al., 2015; Rahman et al., 2013; Drissia, 2019; Cassalho et al., 2019; Mostofi, et al., 2019). The best model is selected on the basis of different goodness of fit tests such as Anderson Darling (AD) test, Kolmogrove-Smirknove (KS) and L-moments ratio

diagram (LMRD). After finding the best models, the next step is to find the results in the form of Pearson correlation coefficient, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Nash-Sutcliffe coefficient and Index of Agreement (IOA) for observed and predicted values and then compare these results with other techniques.

Results and discussions

Data exploration

The annual peak flows data of Mangla site located at River Jhelum, Pakistan was retrieved from WAPDA and federal flood commission of Pakistan. The peak flows normally occur in monsoon season of Pakistan i.e. from July to September. The basic statistics of the data used in this research is given in *Table 1*.

Table 1. Basic information about all of the sites used in the study

Number of observations	Mean (Cusecs)	Standard deviation (Cusecs)	Coefficient of skewness	Coefficient of variation	Latitude	Longitude	Elevation
$n = 54$	132480	136390	4.2402	1.0295	33.15	73.65	14902.7

Table 1 shows that there are total 54 values of peak flows for Mangla site. Each value represents the maximum value in the whole year. The data of peak flows at Mangla site is positively skewed with the coefficient of skewness as 4.2402. The mean of time series is 132480 cusecs. First estimate of first L-moment is equal to mean of the data. The standard deviation of the data is little bit higher than its mean i.e. 136390 cusecs. The coefficient of variation for this time series is 1.0295.

ANN and W-ANN modeling

Annual peak flows for 54 years (1960-2013) of Mangla site on River Jhelum in Pakistan are used for the performance evaluation of Artificial Neural Network (ANN) and Wavelet-ANN Conjunctural model (W-ANN). Eighty percent annual peak flows (1960-2002) are used for training ANN and Wavelet-ANN conjunctural model. The remaining twenty percent data (2003-2013) are used for testing the trained model,

By using the Discreet Wavelet-transform (DWT-Haar wavelet level 1), the approximation and details sub-signals are obtained from the temporal characteristics of of annual peak observed annual peak flows, which are considered as the inputs for W-ANN model. The output for W-ANN model is the original peak flows in the present time step (*Fig. 5*). In the previous step, original values flows are input and in the present time step the values of annual peak flows are used as output to the T-ANN model. In this way, for both models the output is same.

The approximation and detail sub-signals of the annual peak flows series by using DWT db1 (Haar wavelet level 1) are shown in *Figure 6a* and *b*. The approximation and detail sub-signals for DWT db2 (Haar wavelet level 2) are shown in *Figure 7a* and *b*. In this study, two W-ANN models are introduced for comparing purpose such as db1 and db2 respectively. Utilizing db1 transform as inputs for the first W-ANN model and db2 transform as inputs for the second W-ANN model.

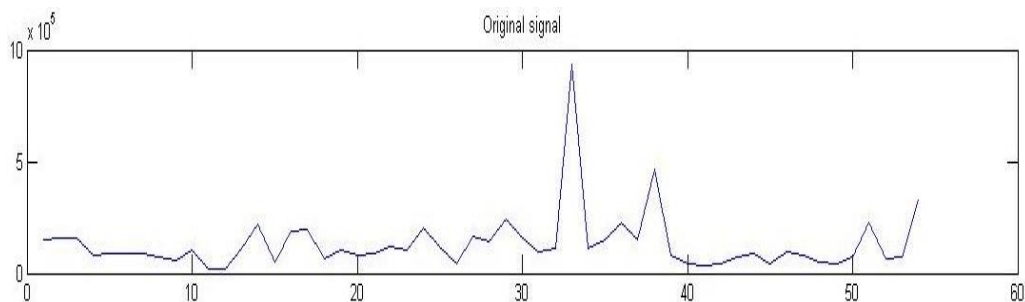


Figure 5. Plot of Annual peak flows (Cusecs) at Mangla site

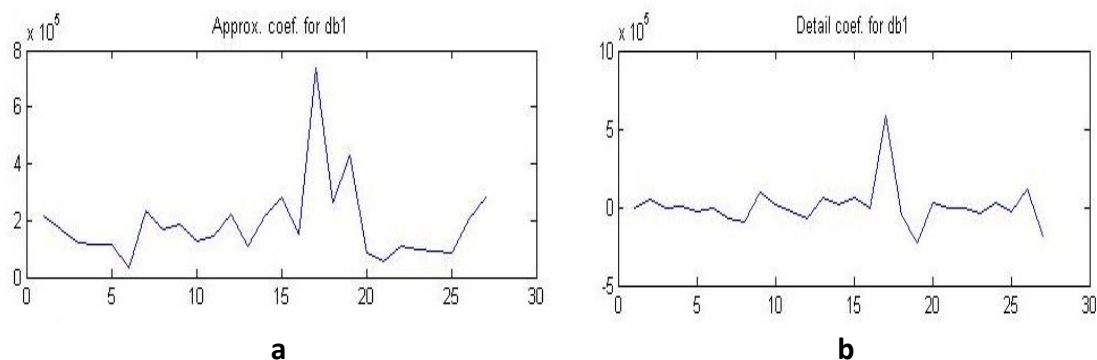


Figure 6. a Approximation sub-signal of peak flows by DWT db1 at Mangla site. **b** Detail sub-signal of peak flows by DWT db1 at Mangla site

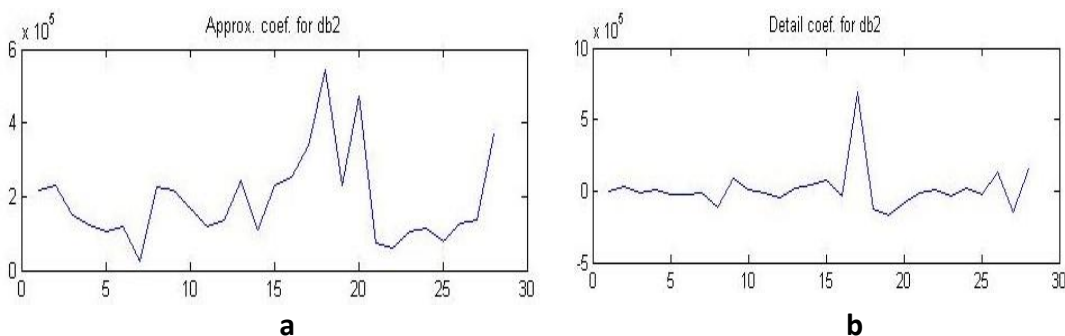


Figure 7. a Approximation sub-signal of peak flows (cusecs) by DWT db2 at Mangla site. **b** Detail sub-signal of peak flows (cusecs) by DWT db2 at Mangla site

By using the methodology shown in *Figure 4*, the (W-ANN) conjunctional model is used here. Back propagation neural network training algorithm is executed on MATLAB platform. Internal parameters used are, momentum coefficient = 0.5 and number of iterations = 1000 to obtain the ideal architecture. Furthermore, ideal combinations of activation functions in both hidden and output layers are obtained with sigmoid function. In this study we consider two input, two hidden, and one output node (2-2-1) for W-ANN conjunctional model and one input, two hidden, and one output nodes (1-2-1) for T-ANN model. The performance of both models W-ANN-db1 (2-2-1) and T-ANN model (1-2-1) for one step ahead forecasting is compared in *Table 2*.

Table 2. Training and testing errors for T-ANN and W-ANN-db1

Models	Evaluation Measures					
	Training/testing	R	RMSE (Cusecs)	E	IOA	MAE (Cusecs)
W-ANN (2-2-1)	Training	0.6394	1.63E+04	0.4914	0.8965	1.83E+04
	Testing	0.5631	3.85E+04	0.3085	0.7382	2.96E+04
T-ANN (1-2-1)	Training	0.4963	2.13E+04	0.2168	0.7317	1.96E+04
	Testing	0.4891	4.26E+04	0.0938	0.5861	3.05E+04

Table 2 shows the performance of W-ANN and T-ANN model in both phases training and testing by using various statistical evaluations measures. W-ANN-db1 (2-2-1) and T-ANN (1-2-1) are chosen as best models among class of other W-ANN and T-ANN models. From Table 2, it is observed that prediction for testing data is not good as compared to the training data using evaluation measures for W-ANN-db1 and T-ANN. From Table 2, the R (0.6394 and 0.5631) for training and testing data of W-ANN-db1 model is greater than the R (0.4963 and 0.4891) for training and testing data respectively of T-ANN model. The RMSE of W-ANN-db1 model for training and testing are (1.63E+04 and 3.85E+04) which is less than the RMSE (2.13E+04 and 4.26E+04) of T-ANN model for training and testing data respectively. Nash-Sutcliffe coefficient E (0.4914 and 0.3085) and Index of Agreement IOA (0.8965 and 0.7382) of W-ANN-db1 model for both training and testing data are also greater than the corresponding E (0.2168 and 0.0938) and IOA (0.7317 and 0.5861) of T-ANN model for training and testing data respectively. The MAE (1.83E+04 and 2.96E+04) for training and testing data of W-ANN-db1 model are less than the MAE (1.96E+04 and 3.05E+04) for training and testing data of T-ANN model respectively. Therefore, due to goodness of fit tests (R, RMSE, E, IOA, & MAE) W-ANN conjunctive model (db1) with two input, two hidden, and one output nodes i.e. W-ANN-db1 (2-2-1) is selected as best performing model as compare T-ANN i.e. T-ANN (1-2-1).

Also, Figure 8, given below compares the actual and predicted peak flows in result of T-ANN (1-2-1) model. Peak flows in the start were accurately forecasted while at the end these flows were not showing harmony with each other. Figure 9 compares the actual flows and predicted peak flows in the result of W-ANN-db1 (2-2-1) model. Peak flows were forecasted in a better way at different parts as compared to previous T-ANN model.

Results are further improved when approximation and detail sub-signal of the annual peak flows are obtained using DWT at db2 (Haar wavelet level 2) as presented in Table 3. Results are improved with W-ANN-db2 (2-2-1) model as compared to W-ANN-db1 (2-2-1). Table 3, shows the results of R, RMSE, E, IOA and MAE for T-ANN (1-2-1) and W-ANN-db2 (2-2-1) model for Mangla site.

Table 3 shows the results of W-ANN-db2 model for Mangla site which clearly shows that the R is improved from 0.6394 to 0.7984 for training data and from 0.4963 to 0.6108 for testing data. The value RMSE are reduced from 1.63E+04 to 1.21E+04 for training and from 3.85E+04 to 2.99E+04 for testing data. The statistical measure Nash-Sutcliffe coefficient E is also improved from 0.4914 to 0.6362 and from 0.3085 to 0.4158 for training and testing data respectively. Similarly, the term Index of Agreement (IOA) is improved from 0.8965 to 0.9135 and from 0.7382 to 0.7916 for

training and testing data respectively. MAE are reduced from 1.83E+04 to 1.04E+04 and from 2.96E+04 to 2.11E+04 for training and testing data respectively. It could be observed from *Table 3* that the performance of W-ANN-db2 model is better than the W-ANN-db1 model for Mangla station.

Table 3. Training and testing errors for T-ANN and W-ANN-db2

Models	Evaluation measures					
	Training/ Testing	R	RMSE (Cusecs)	E	IOA	MAE (Cusecs)
W-ANN (2-2-1)	Training	0.7984	1.21E+04	0.6362	0.9135	1.04E+04
	Testing	0.6108	2.99E+04	0.4158	0.7916	2.11E+04
T-ANN (1-2-1)	Training	0.4963	2.13E+04	0.2168	0.7317	1.96E+0R
	Testing	0.4891	4.26E+04	0.0938	0.5861	3.05E+04

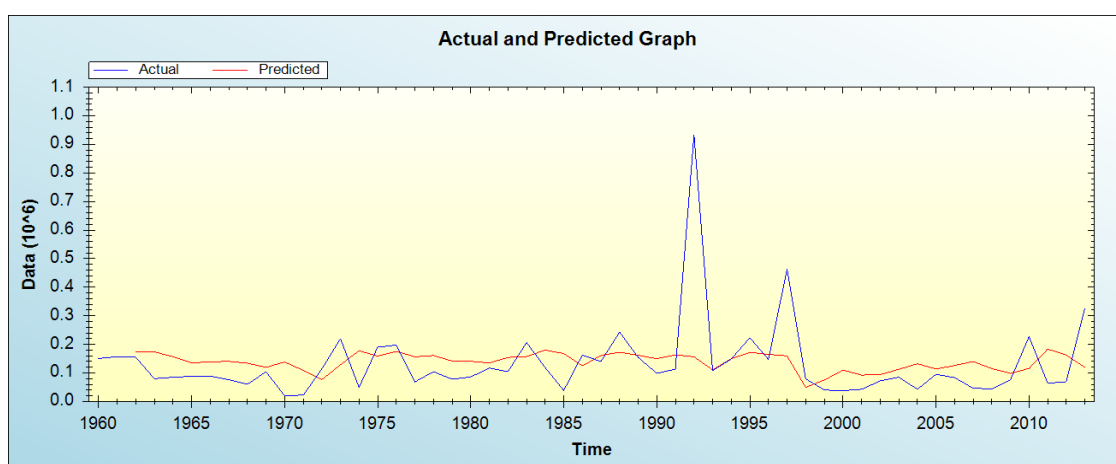


Figure 8. Graph of observed and predicted flows (cusecs) using T-ANN-db1(1-2-1) for Mangla site

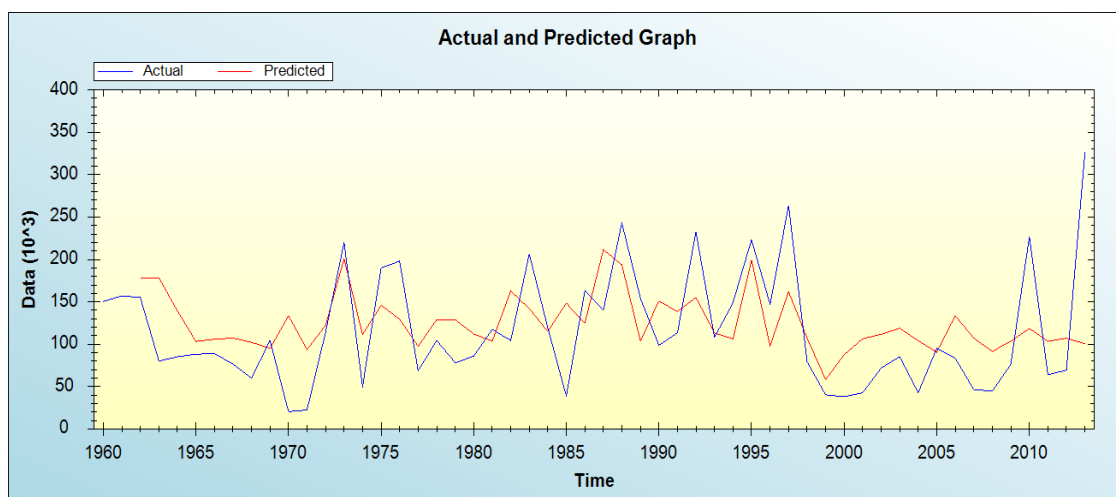


Figure 9. Graph of observed and predicted flows (cusecs) using W-ANN-db1(2-2-1) for Mangla site

Figure 10 compares the actual and predicted peak flows from W-ANN-db2 model. Peak flows at different stages were accurately forecasted. It can be seen that the W-ANN-db1 model was not suitable model for forecasting purpose in the present study. On the other side, W-ANN-db2 model provides closer estimates of predicted values to the corresponding observed peak flows for a lead-time of 1 year.

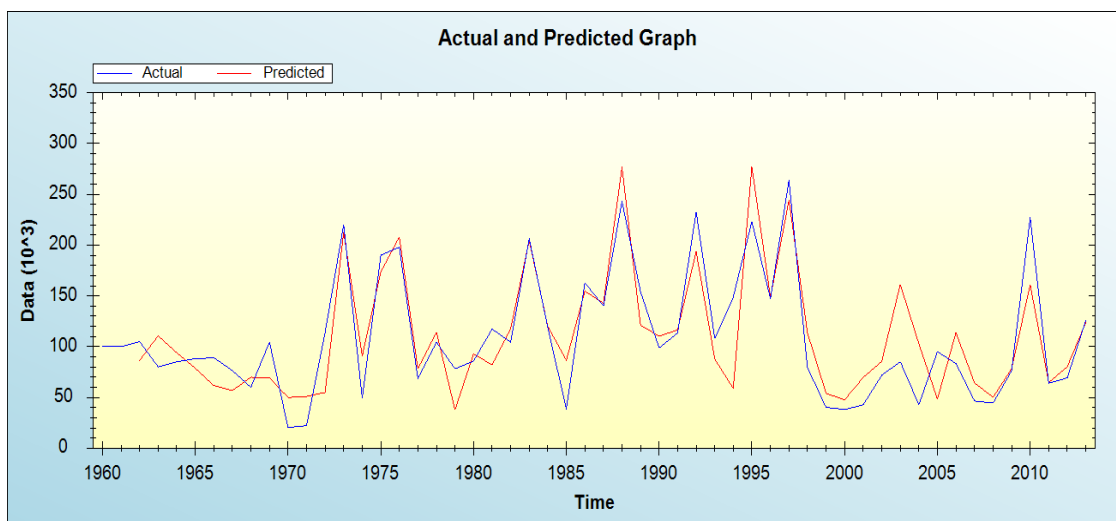


Figure 10. Graph of observed and predicted flows (cusecs) using W-ANN-db2(2-2-1) for Mangla site

Probability distribution modeling and kernel density estimation of annual peak flows

There are four basic assumptions of a time series which must be checked before any further probability distribution modeling in statistical hydrology. These assumptions are randomness, stationarity, independence, and homogeneity. These assumptions are of great importance, without them the results are not accurate and may affect the policy implications. We tested these assumptions with the help of different tests such as Mann Kendall's test for trend detection and Mann-Whitney test for homogeneity with respect to location parameters, and for independence we applied the Lag1 correlation test. One sample runs test is also applied for detection of randomness in the present study.

Table 4 shows that all of the fundamental assumptions are fulfilled and we can say that the time series of peak flows is suitable for further analysis. In the next step, we will find best suitable model among GLO, GPA, GEV, GNO and PE3. We used different goodness of fit tests such as Anderson – Darling (AD), Kolmogrove-Smirknove (KS) test and L-moments ratio diagram (LMRD). For estimation of parameters of these distributions we used method of linear moments (L-moments) because of its outperformance in the presence of extreme observations. From KS and AD test, GLO got first rank and declared as the most suitable probability model for peak flows of Mangla site. The same data has also been used in Ahmad et al. (2015). LMRD is a visual inspection for selecting the most appropriate probability model among class of different probability models. It is a graph of theoretical L-moments ratio of skewness and kurtosis of different probability models. This ratio is represented by a curve in the case of three parameters distribution. The sample ratio of these L-moments ratios using the data of Mangla site is shown with “+” sign. The closeness of this “+” to

any distribution's curve will identify the suitability of that probability model for a particular data. This LMRD, in the present case, also suggest that GLO distribution is the most suitable model for Peak flows of Mangla site see *Figure 10*. The result of LMRD and other goodness fit tests show the same result. After selection of the most suitable model, the next step is find different evaluation measures such as R, RMSE, IOA, MAE for comparison purposes with other approaches. The results are given in *Table 5*.

Table 4. Results of different assumption for annual peak flows (cusecs) at Mangla site

Assumption of stationarity		Assumption of homogeneity		Assumption of independence		Assumption of randomness	
Test statistic	P-value	Test statistic	P-value	Test statistic	P-value	Test statistic	P-value
-0.154	0.238	69	0.074	0.236	0.174	14	1

Table 5. Results of various evaluation measures through probability distribution modeling and kernel density estimation

Models	Evaluation measures				
	R	RMSE (cusecs)	E	IOA	MAE (cusecs)
Probability distribution modeling	0.9401	1.94E+04	0.3513	0.8661	1.55E+04
Kernel density estimation	0.7415	3.8E+04	0.5035	0.5940	5.01E+04

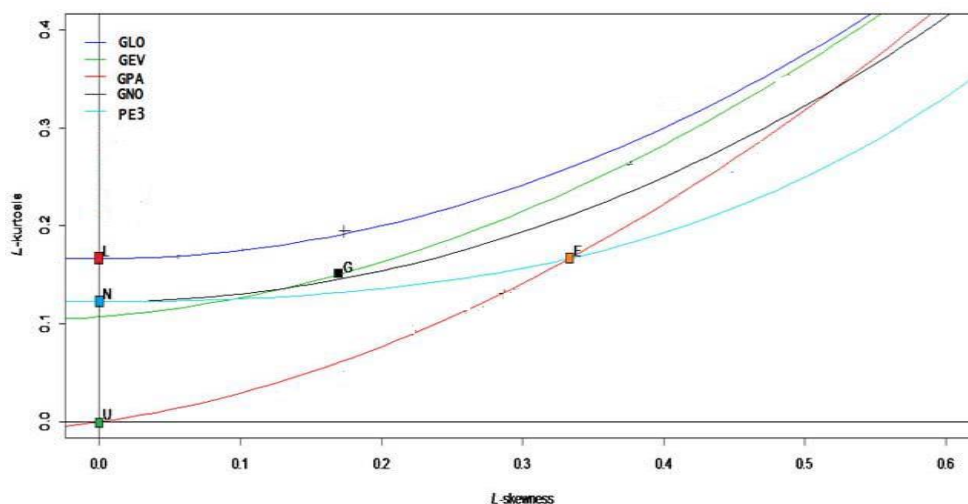


Figure 10. L-moments ratio diagram for Mangla site

For KDE, we use Gaussian kernel and cross validation technique for estimation of smoothing parameter. The choice of kernel is not important as compared to choice of smoothing parameter. In KDE, there is no need to check the assumptions of data like probability distribution modeling. The results of KDE are also presented in *Table 5*.

Results show that probability distribution modeling approach is providing comparable results with W-ANN-db2 model, but outperforms than KDE. Overall the performance of W-ANN-db2 model is better than all other approaches discussed in this study for peak flows of Mangla site.

The results of non-parametric KDE are also presented in the *Table 5*, which are not so promising as W-ANN-db2 and Probability distribution modeling approach. Over all, the W-ANN-db2 approach outperforms than T-ANN, probability modeling and KDE.

Conclusion

In this study the annual peak flows series measured in cusecs at Mangla site situated on the river Jhelum, Pakistan with length of 54 years has been considered for identification of the most suitable modeling approach among T-ANN, W-ANN, Probability distribution modeling and non-parametric KDE. In the light of statistical evaluation criteria, it can be seen that W-ANN model under db1 gave more accurate forecasting results than the T-ANN model. The results further are improved by fitting the W-ANN conjunctive models under db2 scheme. Further to check the authenticity of W-ANN-db2 model, we compared it with probability distribution modeling approach by fitting GLO probability model being the most suitable model and KDE approach using Gaussian kernel approach. The comparison reveals that, probability distribution modeling approach gives us comparable results as of W-ANN-db2 model. The results of KDE are not as promising as of W-ANN and probability distribution modeling approach. Overall, comparison among four approaches, we notice that W-ANN-db2 model outperformed in this study. These results are useful in water resources management such as construction of different hydrological structures.

Acknowledgements. Authors are very grateful to deanship of scientific research at King Khalid University, Abha, Saudi Arabia for the financial support through General Research Program under project number GRP-133-40.

REFERENCES

- [1] Agarwal, A., Singh, R. D., Mishra, S. K., Bhunya, P. K. (2005): ANN-based sediment yield models for Vamsadhara River basin (India). – *Water Sa* 31(1): 85-100.
- [2] Ahmad, I., Fawad, M., Mahmood, I. (2015): At-site flood frequency analysis of annual maximum stream flows in Pakistan using robust estimation methods. – *Polish Journal of Environmental Studies* 24(6): 11-20.
- [3] Aziz, K., Rahman, A., Shamseldin, A. (2016): Development of Artificial Intelligence Based Regional Flood Estimation Techniques for Eastern Australia. – In: Shanmuganathan, S., Samarasinghe, S. (eds.) *Artificial Neural Network Modelling*. Springer, Cham.
- [4] Beven, K. J. (2011): *Rainfall-Runoff Modelling: The Primer*. – John Wiley & Sons, Lancaster, UK.
- [5] Bishop, C. M. (1996): *Neural Networks for Pattern Recognition*. – Oxford University Press, New York.
- [6] Cassalho, F., Beskow, S., de Mello, C. R., de Moura, M. M., de Oliveira, L. F., de Aguiar, M. S. (2019): Artificial intelligence for identifying hydrologically homogeneous regions: A state-of-the-art regional flood frequency analysis. – *Hydrological Processes*. <https://doi.org/10.1002/hyp.13388>.

- [7] Dawson, C. W., Wilby, R. L. (2001): Hydrological modelling using artificial neural networks. – *Progress in Physical Geography* 25(1): 80-108.
- [8] Denis, C., Assis, W. O., Coelho, A. D. (2009): Sorting tomatoes for industrial processing through of computer vision system based on neural networks. – COBEM 2009– International Congress of Mechanical Engineering, Gramado, RS.
- [9] Derks, E. P. P. A., Pastor, M. S., Buydens, L. M. C. (1995): Robustness analysis of radial base function and multi-layered feed-forward neural network models. – *Chemometrics and Intelligent Laboratory Systems* 28(1): 49-60.
- [10] Drissia, T. K., Jothiprakash, V., Anitha, A. B. (2019): Flood frequency analysis using L moments: a comparison between at-site and regional approach. – *Water Resources Management* 2019: 1-25.
- [11] Gramacki, A. (2018): Nonparametric Kernel Density Estimation and Its Computational Aspects. – Springer International Publishing, New York.
- [12] Grossmann, I. E., Guillén-Gosálbez, G. (2010): Scope for the application of mathematical programming techniques in the synthesis and planning of sustainable processes. – *Computers & Chemical Engineering* 34(9): 1365-1376.
- [13] Guresen, E., Kayakutlu, G., Daim, T. U. (2011): Using artificial neural network models in stock market index prediction. – *Expert Systems with Applications* 38(8): 10389-10397.
- [14] Hertz, J. A. (2018): Introduction to the Theory of Neural Computation. – CRC Press, Boca Raton, FL.
- [15] Hipel, K. W., McLeod, A. I. (1994): Time Series Modelling of Water Resources and Environmental Systems. (Vol. 45). – Elsevier, Amsterdam.
- [16] Huang, G., Huang, G. B., Song, S., You, K. (2015): Trends in extreme learning machines: A review. – *Neural Networks* 61(1): 32-48.
- [17] Jimeno-Sáez, P., Senent-Aparicio, J., Pérez-Sánchez, J., Pulido-Velazquez, D., Cecilia, J. (2017): Estimation of instantaneous peak flow using machine-learning models and empirical formula in peninsular Spain. – *Water* 9(5): 347-359.
- [18] Khosrow-Pour, M. (ed.) (2006): Dictionary of Information Science and Technology. (Vol. 1). – IGI Global, Hershey, Pennsylvania.
- [19] Kisi, O., Shiri, J. (2011): Precipitation forecasting using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models. – *Water Resources Management* 25(13): 3135-3152.
- [20] Kumar, R., Goel, N. K., Chatterjee, C., Nayak, P. C. (2015): Regional flood frequency analysis using soft computing techniques. – *Water Resources Management* 29(6): 1965-1978.
- [21] Lall, U. (1995): Recent advances in nonparametric function estimation. Hydrologic applications. – *Reviews of Geophysics* 33(S2): 1093-1102.
- [22] Maidment, D. R. (1993): Handbook of Hydrology. – McGraw-Hill, New York.
- [23] Minns, A. W., Hall, M. J. (1996): Artificial neural networks as rainfall-runoff models. – *Hydrological Sciences Journal* 41(3): 399-417.
- [24] Mostofi Zadeh, S., Durocher, M., Burn, D. H., Ashkar, F. (2019): Pooled flood frequency analysis: a comparison based on peaks-over-threshold and annual maximum series. – *Hydrological Sciences Journal*. <https://doi.org/10.1080/02626667.2019.1577556>.
- [25] Mukerji, A., Chatterjee, C., Raghuvanshi, N. S. (2009): Flood forecasting using ANN, neuro-fuzzy, and neuro-GA models. – *Journal of Hydrologic Engineering* 14(6): 647-652.
- [26] Ogden, T. (2012): Essential Wavelets for Statistical Applications and Data Analysis. – Springer Science & Business Media, New York.
- [27] Rahimi-Ajdadi, F., Abbaspour-Gilandeh, Y. (2011): Artificial neural network and stepwise multiple range regression methods for prediction of tractor fuel consumption. – *Measurement* 44(10): 2104-2111.
- [28] Rahman, A. S., Rahman, A., Zaman, M. A., Haddad, K., Ahsan, A., Imteaz, M. (2013): A study on selection of probability distributions for at-site flood frequency analysis in Australia. – *Natural Hazards* 69(3): 1803-1813.

- [29] Rahman, A., Charron, C., Ouarda, T. B., Chebana, F. (2018): Development of regional flood frequency analysis techniques using generalized additive models for Australia. – *Stochastic Environmental Research and Risk Assessment* 32(1): 123-139.
- [30] Requena, A. I., Ouarda, T. B., Chebana, F.(2018): Low-flow frequency analysis at ungauged sites based on regionally estimated streamflows. – *Journal of Hydrology* 563: 523-532.
- [31] Salas, J. D. (1980): *Applied Modeling of Hydrologic Time Series*. – Water Resources Publication, Littleton.
- [32] Shabri, A. (2002): Nonparametric kernel estimation of annual maximum stream flow quantiles. – *Matematika* 18(1): 99-107.
- [33] Silverman, B. W. (2018): *Density Estimation for Statistics and Data Analysis*. – Routledge, Boca Raton, FL.
- [34] Singh, R. M., Datta, B., Jain, A. (2004): Identification of unknown groundwater pollution sources using artificial neural networks. – *Journal of Water Resources Planning and Management* 130(6): 506-514.
- [35] Tiwari, H., Rai, S. P., Sharma, N., Kumar, D. (2015): Computational approaches for annual maximum river flow series. – *Ain Shams Engineering Journal*. DOI: 10.1016/j.asej.2015.07.016.
- [36] Willmott, C. J., Robeson, S. M., Matsuura, K. (2012): A refined index of model performance. – *International Journal of Climatology* 32(13): 2088-2094.
- [37] Zurada, J. M., Malinowski, A., Usui, S. (1997): Perturbation method for deleting redundant inputs of perceptron networks. – *Neurocomputing* 14(2): 177-193.