A COMPARISON OF QUADRATIC REGRESSION AND ARTIFICIAL NEURAL NETWORKS FOR THE ESTIMATION OF QUANTILES AT UNGAUGED SITES IN REGIONAL FREQUENCY ANALYSIS

KHAN, M. S. R.[1*] – HUSSAIN, Z.[2] – AHMAD, I.[3,1]

[1]*Department of Mathematics and Statistics, International Islamic University
H-10 Islamabad, Pakistan*

[2]*Research Center for Modeling and Simulation (RCMS), National University of Sciences and Technology (NUST), H-12 campus Islamabad, Pakistan*

[3]*Department of Mathematics,College of Sciences, King Khalid University, 61413 Abha, Saudi Arabia*

[*]*Corresponding author
e-mail: shafiqnaizi@gmail.com (phone: +92-334-760-3022)*

Abstract. The study illustrates application of Regional Flood Frequency Analysis (RFFA) using Annual Maximum Peak Flows (AMPF) of eleven gauging sites of various streams of Khyber-Pakhtunkhwa, Pakistan. Assumptions associated to recorded data at various sites have been validated through various statistical tests. The discordancy measure indicates that there is no discordant site in the cluster of eleven sites. Heterogeneity measure based on l-moments confirms that the group of eleven sites is definitely homogeneous. Criterion of $|Z - Dist|$ statistic and L-moment ratio diagram show that Generalized Pareto (GPA) distribution is the best fitted regional distribution of the study region. Regional flood quantiles for various return periods have been estimated using the quantile function of GPA distribution. Artificial Neural Networks (ANN) and Quadratic Regression (QR) model with robust estimation method have been used for the estimation of quantiles at ungauged sites. Model evaluation criteria's (error comparison of predicted values) suggested that estimated quantiles through ANN are accurate relative to quadratic regression. Historical comparison shows that the quantiles estimated through index flood method and ANN are closely related to the highest recorded values of AMPF at each corresponding site for shorter as well as longer return periods.
Keywords: *analyzing extremes of floods, L-moments, regional frequency analysis, Khyber-Pakhtunkhwa, Pakistan, machine learning, non-linear regression*

Abbreviations: AMPF, Annual Maximum Peak flows; RFFA, Regional Flood Frequency Analysis; KPK, Khyber Pakhtunkhwa; AARF, Average Anuual Rainfall; ARMS, Average Rainfall in Monsoon; long, Longitude; lat, Latitude; ele, Elevation; QR, Quadratic Regression; ANN, Artificial Neural Networks; RMSE, Root Mean Square Error; MAPE, Mean Absolute Percentage Error; MAE, Mean Absolute Error; GPA, Genralized Pareto Distribution; GLO, Genralized Logistic Distribution; GNO, Genralized Normal Distribution; GEV, Genralized Exterem Value Distribution; PE3, Pearson Type Three Distribution; l_1, First sample L-moment

Introduction

Reliable estimation of extreme hydrological events is needed for designing and building of hydrological structures on stream channels. These structures are important to provide protection against floods and to regulate the supply of available water. Various approaches including at-site and regional are in practice for flood frequency analysis in different parts of the world. However, regional approach has advantageous results relative to at-site analysis. For instance, the estimated flood quantiles can be

interpolated and/or extrapolated to obtain estimates for the poorly gauged or ungauged sites of the study area. The most commonly used techniques for the estimation of quantiles at ungauged sites include the regression techniques, rational method, ANN, etc. The current study has adopted RFFA using index flood method based on L-moments coupled with QR with robust estimation method and ANN, to estimate flood quantiles at gauged and ungauged or poorly gauged sites, emphasizing the complete justifications related to the regression modeling and comparison with the historical record to investigate the practical validity of the resultant estimates.

Pakistan, with integrated river basins, has a long history of floods since 1947 (from the year of its independence). Twenty-four major floods have occurred in the country from 1947 to 2016, resultantly, the country has suffered direct economic loss of about 38.171 million USD (Government of Pakistan, 2016). Secondly, Pakistan, being a developing country with agriculture sector as a major contributor to its gross domestic product (GDP) requires effective planning and management, based on reliable estimates, for optimum utilization of the available water resources. Agriculture sector contributes about 24 percent in the GDP of Pakistan (Pakistan Bureau of Statistics, 2018). Thirdly, Pakistan's major rivers originate in India and Jammu & Kashmir (a disputed territory between India and Pakistan), and flows from north to south in the country. Due to the Indus Waters Treaty signed in 1960 between India and Pakistan for the division of major rivers of Indus river basins, and the lack of planning and development of the hydrological structures (dams and barrages) on the major rivers (Jhelum, Chenab and Indus), Pakistan is facing problems of shortage of water every year. This crisis will become even worst in future (Development Advocate Pakistan, 2017), as officials are forecasting swear crises of usable water by the year 2025. Therefore, planning, management and efficient utilization of available water resources, especially small rivers and stream flows of north-western areas of Pakistan, which originates in Pakistan, using standard methods of analysis is a primary need of time. The current study has focused, for the first time (to the best of author's knowledge), on various sites of the area of Khyber Pakhtunkhwa (KPK) (an important province in the north-west region of the country) to perform RFFA.

RFFA basically involves two principle steps, (1) identification of regions having similar site characteristics and (2) development of forecast equations for the estimation of flood quantiles at gauged and ungauged sites. RFFA has widely been used all over the world (including the neighboring countries of Pakistan) for the estimation of extreme events. For example; in Canada, (Glaves and Waylen, 1997; Pokhrel, 2002; Requena et al., 2017); in India, (Kumar et al., 2003a,b; Kumar and Chatterjee, 2006; Alam et al., 2016); in China, (Jingyi and Hall, 2004; Yang et al., 2010); in Iran, (Modarres, 2008; Malekinezhad and Zare-Garizi, 2014); in Turkey, (Saf, 2009; Aydoğan et al., 2016); in Italy, (Noto and La Loggia, 2009) and many more. In RFFA, several methods are in practice including regression techniques, rational methods, ANNs, etc. for the development of forecast equations to estimate flood quantiles at ungauged sites. Few details of the adopted methods in published studies are illustrated in the following sections.

Regression techniques have been used by various published studies for the estimation of quantiles at ungauged sites (Griffis and Stedinger, 2007; Hailegeorgis and Alfredsen, 2017), but very few of them have illustrated the complete theoretical and statistical justifications of the developed models. Moreover, scarce literature is available analyzing the nonlinear relationships of the site characteristics with observed AMPF of

gauged sites. Importantly, the relationship between hydrological and physio-meteorological variables is usually categorized as strongly nonlinear (Sivakumar and Singh, 2012). Nonlinear relationships may provide more accurate estimates of flood quantiles at ungauged sites as compared to linear relationships (Ouali et al., 2017). Secondly, due to the existence of outliers in the observed AMPF and availability of fewer values for analysis, the Ordinary Least Squares (OLS) estimation method may no longer be feasible. This problem requires robust methods for the estimation. In the current study, a quadratic term of the most influential site characteristic as independent variable has been introduced in the model along with a robust method for the estimation of the proposed model.

ANN is a nonparametric calculating and modeling approach stimulated by the biological operative of the human brain (Rumelhart et al., 1985). This method has significant advantages over other estimation techniques including regression analyses (Liu et al., 2009; Landi et al., 2010). One major advantage of ANN is their capacity to identify the complex nonlinear relationships and there is no need to express such relationship in mathematical form as the data itself recognizes the model form through use of artificial intelligence (Hjelmfelt and Wang, 1996). ANN approach coupled with RFFA can provide more accurate estimates of flood quantiles for ungauged sites relative to regression models (Dawson et al., 2006; Shu and Ouarda, 2007; Aziz et al., 2014; Anilan et al., 2016). Therefore, adopting recent trends in statistical hydrology, ANN has been used, for the first time in Pakistan, for the estimation of flood quantiles at ungauged sites.

Since, one of the emphases of this study is the identification of optimum method coupled with RFFA to obtain quantiles for ungauged sites. Therefore, the estimates through QR and ANN have been compared using Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE) of predicted values from both the methods of estimation. Moreover, a comparison with the historical flood information has been done to investigate the practical validity of the provided estimates through different methods.

Keeping in view the aforementioned details, major objectives of the study are:
- Estimation of flood quantiles using RFFA based on L-moments for the gauged sites of important streams of KPK with complete justifications of the adopted procedure through various statistical tests.
- Estimating quadratic regression model with robust method of estimation for the estimation of quantiles at ungauged sites.
- Introduction of ANN for the estimation of quantiles for ungagued sites of the study area.
- Comparison of the esitmates obtained through various procedures of estimation for ungauged sites using statistical measures along with histrorical information to identify the optimum method to be used in RFFA for the estimation of quantiles at ungauged sites.

Study area and data utilized

The KPK region, having various small rivers and stream flows, is the second source of river water in Pakistan. KPK has 101,741 km² area with steep geography and a population of about 35.53 million (as per population censes of 2017 by the Government of Pakistan). The terrain of KPK consists of mountainous in the north, sub mountainous and lands surrounded by hills to the south. Due to steep geography and mountain land

of KPK the heavy rainfall usually turns into flash flooding affecting the whole of KPK (Pakistan Meteorological Department, 2012). Southern KPK is the most populated area of the province, and due to its downstream location, it has been affected badly due to heavy floods in 1992 and 2010 (Hashmi et al., 2012). Therefore, preventive measures against these natural disasters are a popular demand of the people of KPK which requires quantification of the frequency associated to these floods. Moreover, KPK has an identified potential of hydro-electricity of about 18698 Mega Watt, as reported in a study on Hydel Potential in Pakistan by National Electric Power Regulatory Authority (NEPRA), Pakistan (NEPRA, 2018). None of the published studies so far has used L-moment based RFFA for flood quantiles estimation at various sites of KPK. Few published studies in Pakistan related to statistical hydrology include (Muhammad and Afreen, 2007; Hussain and Pasha, 2009; Hussain, 2011; Ahmad et al., 2015; Ahmad et al., 2016; Batool, 2017; Hussain, 2017; Hussain et al., 2017 etc.). These studies have focused on the areas of Punjab and Sindh only (the two provinces of Pakistan). Other studies including (Shahzadi et al., 2013; Khan et al., 2017) performed rainfall frequency analysis considering few areas of KPK and Sindh.

The current study has performed RFFA using AMPF in cusecs of eleven gauging sites situated in the southern KPK. The data has been provided by the Provincial Irrigation Department of KPK. *Fig. 1* illustrates the geographical information of the sites used for the analyses. The details of these sites and their respective site characteristics, such as, Latitude (lat), Longitude (long), Elevation (ele), Average Annual Rainfall (AARF) and Average Rainfall in Monsoon (ARMS) are given in *Table 1*. Few missing observations have been found in the observed data series of the sites, Wazir Ghari, Chinkar, Bara Tarnab, Khuderzai, Jundi Utmanzai, Lund Khwar East, Jundi Tangi, and Swat Ningolai. The percentage of these missing observations is available in *Table 1*. These missing observations have been replaced by the average value of observed AMPF of the corresponding site.

Table 1. *Site characteristics of eleven sites of Khyber-Pakhtunkhwa, Pakistan*

S. No.	Site Name	Latitude (North)	Longitude (East)	Elevation (meters)	AARF (mm)	ARMS (mm)	Percentage of missing observations
1	Kalpani Deheri	33.9931	71.7461	303	559	255	0
2	Wazir Ghari	33.9845	71.7749	303	400	119	6
3	Chinkar	34.014	71.754	301	400	119	7
4	Bara Tarnab	34.0168	71.7037	305	400	119	7
5	Khuderzai	34.0121	71.7743	300	532	240	3
6	Jundi Utmanzai	34.0094	71.8328	294	460	170	16
7	Lund Khwar East	34.0060	71.9777	285	559	255	14
8	Kalpani Saidabad	34.0516	71.5282	314	559	255	0
9	Jundi Tangi	34.2826	71.6828	335	460	170	12
10	Swat Ningolai	33.9043	71.5584	379	400	328	6
11	Bara Kohat Road	33.8638	71.5635	413	400	119	0

Khan et al.: A comparison of quadratic regression and artificial neural networks for the estimation of quantiles at ungauged sites in
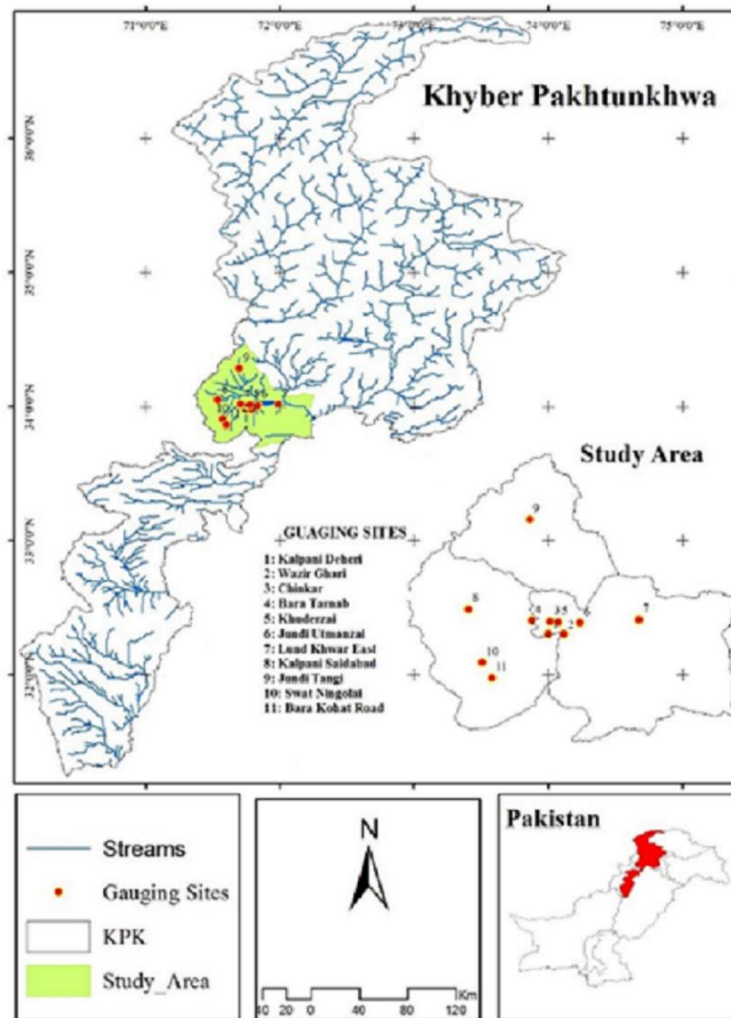regional frequency analysis
- 6941 -

Figure 1. *Study area and geographical locations of eleven gauging sites of Khyber-Pakhtunkhwa, Pakistan*

Methodology

Index Flood Method

This study has adopted L-moments based RFFA, described in details in (Hosking and Wallis, 1997). RFFA is based on certain assumptions related to the observed data series, for instance, the data series is random, stationary and identically distributed etc. These assumptions have been validated using various statistical tests. The first step afterwards is to check whether there exists a discordant site, using Di measure, in a group of available sites. The formula of D_i is:

$$D_i = \frac{1}{3}N(u_i - \bar{u})^T S^{-1}(u_i - \bar{u}), i = 1,2,3 \dots N \qquad \text{(Eq.1)}$$

$$S = \sum_{i=1}^{N}(u_i - \bar{u})(u_i - \bar{u})^T$$

where u_i is the vector of sample L-moments ratios of site i, $\bar{u}$ is its mean and N is the total number of sites. A site is discordant if its D_i value is greater than the critical value. These critical values are provided in (Hosking and Wallis, 1997).

Next important step in RFFA is the identification of homogeneous region. Homogeneity of the identified region can be checked using regional heterogeneity measure H, for which the formula is:

$$H = \frac{V - \mu_v}{\sigma_v} \qquad \text{(Eq.2)}$$

where $V = \left[\frac{\sum_{i=1}^{N} n_i (\tau^i - \tau^R)^2}{\sum_{i=1}^{N} n_i}\right]^{\frac{1}{2}}$ and μ_v is mean and σ_v is standard deviation of computed inter-site variation figured through simulations using Kappa distribution.

For a region to be declared as statistically homogeneous, the calculated value of H should be less than one. After the identification of homogeneous region, the next obvious step is to check the goodness of fit of the candidate distribution(s) using L-moment ration diagram (a graph of sample L-skewness Vs sample L-kurtosis of the sites) and $Z - Dist$ statistic. The formula for $Z - Dist$ statistic is:

$$Z - Dist = \frac{\tau_4^{Dist} - \tau_4^R + \beta_4}{\sigma_4} \qquad \text{(Eq.3)}$$

where τ_4^{Dist} is the L-kurtosis of fitted candidate regional distribution, τ_4^R is the regional L-kurtosis, σ_4 is standard deviation and β_4 is the bias of τ_4^R obtained through simulations. A distribution having minimum $|Z - Dist|$ value can be considered as the best fitted regional probability distribution.

Parameters of best fitted regional distribution can be estimated through method of L-moments and regional quantiles for various return periods can be estimated using the quantile function of the fitted regional distribution. Using the estimated regional quantiles, at-site quantiles can be estimated using:

$$\hat{Q}_i(F) = l_1^{(i)} \hat{q}(F) \qquad \text{(Eq.4)}$$

where $\hat{Q}_i(F)$ is the estimated quantile of site i, $l_1^{(i)}$ is the first sample L-moment of site i and $\hat{q}(F)$ is the estimated regional quantile for any return period.

Operational configuration of ANN

For ungauges quantiles estimation QR and ANN have been used. ANN consist of input layer, hidden layers, neurons in each layer, activation function and output layer. Among the various methods of ANN, multilayer feed forward network has been selected for the current study. Back propagation method has been used for the training of ANN which may also be referred as Back Propagation Neural Networks (BPNN). BPNN is a popular tool and dominates for predictions of river flow at different points and well suited to various types of hydrological modelling especially for non-linear regression operations (Maier and Dandy, 2000; Abrahart et al., 2004). Structure of threelayer feed forward network has been dispalyed in *Fig. 2*.

Khan et al.: A comparison of quadratic regression and artificial neural networks for the estimation of quantiles at ungauged sites in
regional frequency analysis
- 6943 -

In *Fig. 2* circles represents neurons in the layers of ANN structure and arrow lines for connections between neurons in each layer. Weight function has been used as connection parameter and every input of neurons is multiplied by a weight and then the values are pooled to produce one output value. This value is usually produced with certain bias and then operated by activation function. Therefore, output of each neuron stated as below:

$$Z_j = f(X (T_i - b_j)) \tag{Eq.5}$$

where X is the input vector in the first layer and T_i is the weight vector. Its order of weights leading to the neurons given as:

$$T_{mj} = (t_{1j} \dots. t_{ij} \dots. t_{nj})$$

where $j = 1, 2, \dots n$ and m = number of neurons, t_{ij} = joining weight from the i neuron in the earlier layer to the current neuron.

The output of neuron j is Z_j and obtained by the inner product of the function defined in *Equation 5* where b_j represent the value of bias associated with neuron j.
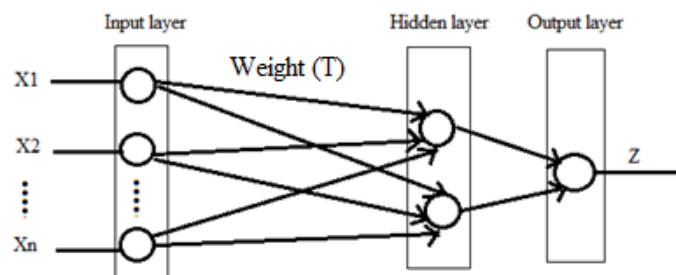

Figure 2. The Configuration of Artificial Neural Networks

The main purpose of training ANN with a data set (input and output) is to adjust the weights to minimize the errors between ANN outputs and the desired outputs. Before training of any neural network, data is scaled using z-normalization, minimum-maximum scale, etc. For the present study minimum-maximum scale has been used to scale the data into the interval [0, 1], and randomly splitting the data into training and testing set. Most commonly known activation functions for neural networks are logistic function, hyperbolic tangent function, bipolar function, scaled arctan function, and arctan function. Shamseldin et al. (2002) states that logistic function performs well relative to other activation function for river flow forecasting. Therefore, logistic function has been used as activation function for hidden and output layers. Number of hidden layers and neurons in each layer has been selected by using trial and error method.

Results and discussion

RFFA is based on few critical assumptions related to the observed data series at each site. For instance, the recorded data at different sites is random, independent and identically distributed as well as free of the step changes and trend. There are various methods available in the literature for the validation of these assumptions. For the current

Khan et al.: A comparison of quadratic regression and artificial neural networks for the estimation of quantiles at ungauged sites in
regional frequency analysis
- 6944 -

study, Run test for randomness (Bradley, 1968), Wald-Wolfowitz test for independence and stationary (Wald and Wolfowitz, 1943), Rank-sum test for homogeneity (Hirsch et al., 1992), have been used and the results are given in *Table 2*. All these tests have been performed using 5% level of significance. The results of *Table 2* show that the recorded data of eleven gauging sites fulfill all the critical assumptions related to RFFA, therefore, can be used for further analysis.

Table 2. *Calculated values of non-parametric test statistics and their corresponding p-values for the validation of assumptions related to the observed data series at various sites*

S.No.	Site name		Rank-Sum	Run Test	Wald-Wolfowitz
1	Kalpani Deheri	Test statistics	0.739	0.459	-0.955
		P-value	0.4599	0.6459	0.3392
2	Wazir Ghari	Test statistics	-0.132	0.574	-0.678
		P-value	0.8950	0.5656	0.4975
3	Chinkar	Test statistics	1.126	-1.201	0.309
		P-value	0.2602	0.2298	0.7573
4	Bara Tarnab	Test statistics	-1.659	-1.858	1.251
		P-value	0.0971	0.0631	0.2109
5	Khuderzai	Test statistics	0.955	-1.274	0.268
		P-value	0.3396	0.2026	0.7887
6	Jundi Utmanzai	Test statistics	0.578	-0.481	1.851
		P-value	0.5633	0.6305	0.0641
7	Lund Khwar East	Test statistics	-1.44	-1.155	0.109
		P-value	0.1499	0.2479	0.9131
8	Kalpani Saidabad	Test statistics	-1.008	-1.797	0.765
		P-value	0.3135	0.0723	0.4443
9	Jundi Tangi	Test statistics	-1.49	1.436	1.182
		P-value	0.1362	0.1510	0.2372
10	Swat Ningolai	Test statistics	1.864	-0.371	1.758
		P-value	0.0623	0.7106	0.0787
11	Bara Kohat Road	Test statistics	-1.171	-0.542	-0.083
		P-value	0.2416	0.5878	0.9339

Discordancy measure

The calculated values of D_i using *Equation1* for eleven sites along with corresponding descriptive statistics in terms of L-moments have been illustrated in *Table 3*. The results show that the values of D_i for each site are less than the critical value i.e. 2.64 (as proposed by Hosking and Wallis (1997) for eleven sites). Therefore, there is no discordant site in the set of eleven gauging sites.

Descriptive statistics given in *Table 3* show that Wazir Ghari, Chinkar and Lund Khwar East have smaller magnitude of AMPF while Bara Tarnab, Kalpani Saidabad and Swat Ningolai have relatively larger magnitude of AMPF in the region. Sample L-moments ratios L-CV (t), L-skewness (t_3) and L-kurtosis (t_4) for the sites Jundi Utmanzai and Jundi Tangi are very high. This shows that there exists high variations in the observed data series at these sites. Moreover, the shape of the distribution associated

to AMPF at these sites is positively skewed with high kurtosis. The sample L-moments ratios for all the sites, in general, show that the observed data series have high variations, positive skewness and hight kurtosis. Moreover, the values of L-skewness are smaller than L-CV while values of L-kurtosis are relatively less than L-skewness for almost all the sites. The site Bara Tarnab have the value of L-skewness greater than L-CV. This high variation within the AMPF may be due to the uncontrolled/natural flow of the stream associated to this site and mountainous and steep geography of the region. An important point is that the flow of the streams in the study region heavily depends on rainfall and the pattern of rainfall in the basin area is irregular. The reason of this high variation within the AMPF has been observed due the presence of outliers, may be due to storm rainfalls in the basin area, especially during monsoon season (from June and August).

***Table 3.** L-moments based descriptive statistics and values of discordancy measures*

S. No.	Site name	n_i	l_1	t	t_3	t_4	D_i
1	Kalpani Deheri	21	2856.61	0.6276	0.4218	0.0894	0.97
2	Wazir Ghari	32	426.47	0.6457	0.5864	0.4113	0.28
3	Chinkar	28	922.27	0.7141	0.6098	0.4117	0.28
4	Bara Tarnab	30	11884.18	0.6480	0.7284	0.6424	2.24
5	Khuderzai	33	1758.37	0.6765	0.5319	0.3616	1.98
6	Jundi Utmanzai	25	2052.57	0.8038	0.7232	0.5031	1.06
7	Lund Khwar East	28	484.08	0.5903	0.4183	0.1166	0.88
8	Kalpani Saidabad	33	9408.82	0.6699	0.5780	0.2949	1.05
9	Jundi Tangi	42	1104.65	0.8156	0.8131	0.6705	1.30
10	Swat Ningolai	33	8933.68	0.6163	0.5349	0.3515	0.53
11	Bara Kohat Road	34	1453.00	0.7243	0.5959	0.3249	0.43

Note: "n_i" is the number of observations at each site, "l_1" is the first sample l-moment, "t" is the sample L-CV, "t_3" is the sample L-skewness, "t_4" is the sample L-kurtosis and "D_i" is the discordancy measure

Regional heterogeneity measure

Initially, the group of eleven sites has been considered as a single region for the analysis. The reason being these sites are geographically contiguous; therefore, their climate conditions, and topography (the variables affecting the flow behavior of the sites under study) may be similar to each other. For the calculations of H, the simulations have been performed using four parameters Kappa distribution. Large number of simulations, i.e. 1000, have been performed to calculate the values of μ_v and σ_v. The calculated value of H by using *Equation 2* is 0.09; therefore, the region consisting of eleven sites is homogeneous and suitable for further analysis.

Fitting of regional probability distribution

After the formation of homogeneous region, the next important step is to identify the best fitted regional probability distribution that accurately defines the observed AMPF within the homogeneous region. Graphical/informal method (L-moment ratio diagram) and formal method ($|Z - Dist|$ statistic) have been used for the selection of best fitted regional distribution for the recorded data series. L-moment ratio diagram is presented in

Khan et al.: A comparison of quadratic regression and artificial neural networks for the estimation of quantiles at ungauged sites in
regional frequency analysis
- 6946 -

Fig. 3, in which sample L-skewness has been plotted against the values of sample L-kurtosis. From *Fig. 3*, it has been observed that the regional average value of sample L-skewness and L-kurtosis lies closer to the theoretical curves of Generalized Normal (GNO) and GPA distributions suggesting their appropriateness as regional distribution. Moreover, the tendency of sample L-skewness and L-kurtosis of all the sites is closer to the theoretical curve of GPA distribution.
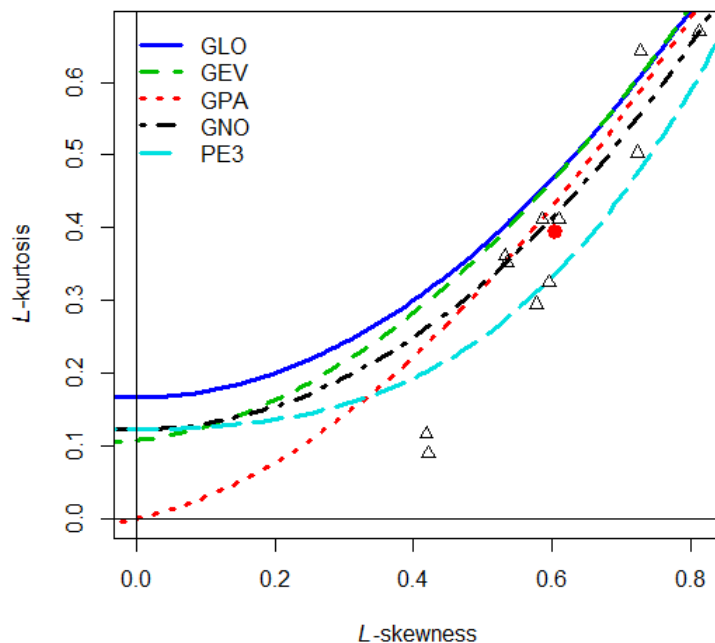

Figure 3. *L-moment ratio diagram. Note: (●) represents the average value of L-skewness & L-kurtosis of the region, (Δ) denotes the values of sample L-skewness & L-kurtosis of each site in the region*

Calculated values of $|Z - Dist|$ statistic, using *Equation 3*, for each candidate distributions i.e. Generalized Extreme Value (GEV), Generalized Logistic (GLO), Pearson type-3 (PE3), GNO and GPA have been given in *Table 4*. Four distributions; GNO, GPA, GEV, and GLO fulfill the criteria for best fitted regional distributions, that is, having value of $|Z - Dist| \leq 1.64$. Among these distributions, GPA distribution has minimum value of $|Z - Dist|$ statistic. Therefore, based on the results of L-moment ratio diagram and $|Z - Dist|$ statistic, GPA distribution has been selected as the best fitted distribution for the region under study.

Table 4. *Values of $|Z - Dist|$ statistic for each candidate regional distribution*

GLO	GEV	GNO	PE3	GPA
0.46	0.39	0.70	2.45	0.34

Note: Genralized Pareto (GPA), Genralized Logistic (GLO), Genralized Normal (GNO), Genralized Exterem Value (GEV), Pearson type-3 (PE3)

Khan et al.: A comparison of quadratic regression and artificial neural networks for the estimation of quantiles at ungauged sites in
regional frequency analysis
- 6947 -

Estimation of regional probability distribution and quantiles

The parameters of GPA distribution have been estimated using method of L-moments. Dimensionless regional quantiles have been estimated using the quantile function of GPA distribution. The estimated parameters and quantiles for different return periods of GPA distribution are given in *Table 5*. The estimated regional quantiles show increasing trend, as usual, for shorter as well as longer return periods.

Table 5. *Estimated parameters of Generalized Pareto (GPA) distribution and regional quantiles for various return periods*

Distribution	Estimated Parameters			Regional quantiles for various return periods in years				
	ε	α	k	10	20	100	500	1000
GPA	-0.0282	0.5064	-0.5075	2.184	3.538	9.303	22.351	32.207

In RFFA, for a homogeneous region, the selected regional distribution is considered as parent distribution for various sites of the region. The estimated at-site quantiles using *Equation 4* for various return periods are given in *Table 6*. The results show that the estimated quantiles at site Wazir Ghari for 100 years return period is 3967 cusecs which is the lowest as compared to other sites in the region. The highest of the estimated quantile for 100 years return period is at site Bara Tarnab as 110560 cusecs. The estimated quantiles at various sites for smaller to larger return periods, i.e. 10, 20, 100, 500 and 1000 years are greater than the average AMPF of their corresponding sites.

Table 6. *Estimated quantiles for each site obtained using index flood method, i.e. Eq. 4, for various return periods*

S. No.	Sites	Quantiles for various return periods in years				
		10	20	100	500	1000
1	Kalpani Deheri	6239	10106	26575	63849	92002
2	Wazir Ghari	931	1508	3967	9532	13735
3	Chinkar	2014	3262	8579	20614	29703
4	Bara Tarnab	25958	42043	110560	265627	382753
5	Khuderzai	3840	6220	16358	39302	56631
6	Jundi Utmanzai	4483	7261	19095	45878	66107
7	Lund Khwar East	1057	1712	4503	10820	15590
8	Kalpani Saidabad	20551	33286	87531	210299	303029
9	Jundi Tangi	2412	3908	10276	24690	35577
10	Swat Ningolai	19513	31605	83110	199679	287727
11	Bara Kohat Road	3173	5140	13517	32476	46796

Quantiles for ungauged sites

For the current study, quadratic regression and ANN have been used to develop a functional relationship between l_1 (mean of observed AMPF at given sites) and their corresponding available site characteristics within the homogeneous region. This functional relationship will provide the estimates of at-site mean (for ungauged sites) within the homogeneous region for the estimation of T-years flood quantiles.

Khan et al.: A comparison of quadratic regression and artificial neural networks for the estimation of quantiles at ungauged sites in
regional frequency analysis
- 6948 -

Quadratic regression analysis

This section describes the development of QR model using the most suitable site characteristic as an explanatory variable (to predict $l_1^{(i)}$, for site i). The details are: *Table 7* provides the correlation matrix, between l_1 and the available site characteristics along with corresponding p-values to check the significance of the respective correlation coefficients.

Table 7. *Values of the correlation coefficients between site characteristics along with their corresponding p-values for testing the significance of correlation coefficients*

Variables		l_1	Latitude	Longitude	Elevation	AARF	ARMS
l_1	r	1	-0.1245	-0.5486	0.1324	-0.0952	0.2656
	(P-value)	(0)	(0.7153)	(0.0805)	(0.7032)	(0.7807)	(0.4299)
Latitude	r	-0.1245	1	0.1377	-0.3690	0.2385	-0.0942
	(P-value)	(0.7153)	(0)	(0.6864)	(0.2641)	(0.4800)	(0.7829)
Longitude	r	-0.5486	0.1377	1	-0.7252	0.2912	-0.0919
	P-value	(0.0805)	(0.6864)	(0)	(0.0116)	(0.3850)	(0.7881)
Elevation	r	0.1324	-0.3690	-0.7252	1	-0.4633	0.0185
	P-value	(0.7032)	(0.2641)	(0.0116)	(0)	(0.1512)	(0.9569)
AARF	r	-0.0952	0.2385	0.2912	-0.4633	1	0.5850
	P-value	(0.7807)	(0.4800)	(0.3850)	(0.1512)	(0)	(0.0587)
ARMS	r	0.2656	-0.0942	-0.0919	0.0185	0.5850	1
	P-value	(0.4299)	(0.7829)	(0.7881)	(0.9569)	(0.0587)	(0)

Note: " r " is the correlation coefficient

The results show that the correlations between l_1 and the available site characteristics are statistically insignificant at 5% level of significance; hence there does not exist significant linear relationship between l_1 and the available site characteristics. Therefore, simple/multiple linear regression model (using linear form of available site characteristic(s)) would not be appropriate to predict the values of l_1. The next obvious choice is the use of quadratic forms of the pertinent site characteristic(s) among available to predict the average value of l_1 for ungauged sites. To select the best independent variable for inclusion in the regression model with linear and quadratic form, following details have been considered. Results of *Table 7* show that the correlation between l_1 and longitude is the highest relative to all other sites characteristics but the problem is that longitude and latitude are imaginary lines on earth glob to identify the geographical location of any site. These site characteristics may be negligible if additional site characteristics are available (Hussain, 2017). The next obvious choice of variable having highest correlation with l_1 as compared to other available site characteristics is ARMS. Therefore, it has been considered as an explanatory variable in the model. Moreover, *Table 8* provides the percentage of occurrence of observed AMPF at different sites during various seasons of a year which shows that a large percentage of values of AMPF have been observed in the monsoon season (from June to August). Therefore, ARMS can be considered as the most influential variable having relatively direct and strong relationship with the observed AMPF in the study region. Based on the aforementioned details, it is

appropriate to use quadratic form of ARMS as explanatory variable in the regression model. Scatter plot of l_1 and ARMS is illustrated in *Fig. 4* in which a nonlinear relationship between the variables is obvious.

Table 8. *Percentage (%) of frequency of AMPF (Annual Maximum Peak Flows) in four seasons at each site*

S. No.	Site name	n_i	Monsoon (%)	Autumn (%)	Winter (%)	Spring (%)
1	Kalpani Deheri	21	85.7	4.7	0.0	9.5
2	Wazir Ghari	32	46.8	12.5	12.5	21.8
3	Chinkar	28	64.2	7.1	10.7	10.7
4	Bara Tarnab	30	56.6	3.3	3.3	30.0
5	Khuderzai	33	51.5	12.1	9.1	24.2
6	Jundi Utmanzai	25	80.0	0.0	0.0	4.0
7	Lund Khwar East	28	64.2	17.8	0.0	3.5
8	Kalpani Saidabad	33	84.8	12.1	3.0	0.0
9	Jundi Tangi	42	69.04	9.5	0.0	9.5
10	Swat Ningolai	33	81.0	10.0	0.0	3.0
11	Bara Kohat Road	34	83.0	5.0	0.0	12.0

Note: The time period for Monsoon is from June to August, autumn is from September to November, winter is from December to February and spring is from March to May
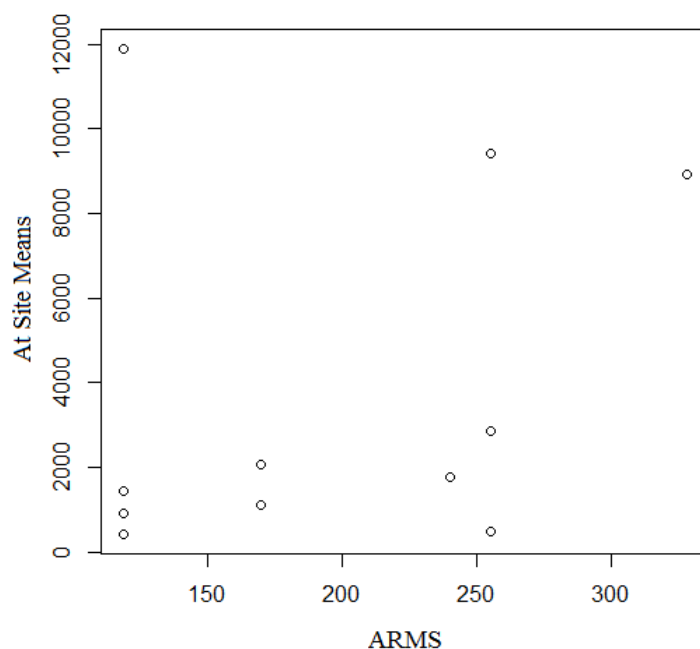

Figure 4. *Scatter plot between l_1 (first sample l-moment of observed annual maximum peak flows) and ARMS (Average rainfall in monsoon)*

Few observations in the data series do not follow the usual pattern of the data and create high scatter in the data. In classical linear regression modeling, such observations

Khan et al.: A comparison of quadratic regression and artificial neural networks for the estimation of quantiles at ungauged sites in
regional frequency analysis
- 6950 -

(outliers) create problems of estimation and in such a situation it is very difficult to fulfill the critical assumptions of classical linear regression (normality and constant variance of the error term).

Therefore, robust estimation methods are obvious choice. For the current study, due to the existence of high scatter in the observed data series and the dependent variable is the mean of extreme values, M-estimation method (Huber, 1973) with covariance type Huber type-III as weight function has been used to estimate the developed regression model. M-estimation technique does not rely on the normality assumption of error term and is more resistant to outliers (Stuart, 2011). M-estimates are consistent, asymptotically unbiased and normally distributed (Chen and Wu, 1989; Hengjian, 1997; Ivanov and Orlovs, 2008; Ruckstuhl, 2014). Thus the fitted regression model based on the aforementioned details is.

$$\widehat{l_1} = 9220.21 - 104.20\ ARMS + 0.3107\ (ARMS)^2 \qquad (Eq.6)$$

Values of the coefficients, standard errors of estimates, t-calculated (to validate the statistical significance of the provided coefficients) and their corresponding p-values are given in *Table 9*.

Table 9. Estimated values of the fitted regression model, coefficients and their corresponding standard errors (S.E.), t-values and P-values

S. No.	Independent variables	Coefficients	S.E.	t-value	P-value
1	ARMS	−104.20	32.01	-3.2554	0.001
2	$(ARMS)^2$	0.3107	0.0719	4.3186	0.000

Note: ARMS (Average rainfall in monsoon)

The results show that the coefficients of the fitted regression model in *Equation 6* are highly significant at 5% level of significance. Quadratic term of the model have positive impact on flood flows when rainfall during monsoon season is increased from its average value. Value of the adjusted R_w^2 is 0.89 for the fitted model. Diagnostics plots of fitted vs residuals and residual vs leverage are given in *Fig. 5* and *Fig. 6*, respectively.

These plots show that high residuals are observed for the 4th site (Bara Tarnab) and 8th site (Kalpani Saidabad) observations. These high residuals lies on the cook's distance limit. Diagnostics plots also show that the 10th observation (the site Swat Ningolai) has deviated largely from other plotted points, as its predicted value is very large and much closer to its observed average AMPF. Therefore, these observations don't have any undue influence on the estimated values. These details reveal that the developed regression model is statistically sound; however, another desirable is the practical validity of the model, which has been provided in second last section.

Artificial neural networks

Five input and one output variables have been used for the prediction of the dependent variable (average value of the observed AMPF at various sites (l_1)). Site characteristics, such as, "lat", "long", "ele", "AARF" and "ARMS" have been used as input variables of ANN model. The variable l_1 has been used as the dependent variable of the model. Functional relationship of the dependent and independent variables is given as

Khan et al.: A comparison of quadratic regression and artificial neural networks for the estimation of quantiles at ungauged sites in
regional frequency analysis
- 6951 -

$$f(l_1) = g(lat, long, ele, AARF, ARMS) \hspace{2cm} \text{(Eq.7)}$$

Two hidden layers, with four neurons in first layer and two in second layer, have been used to perform the algorithm. The graphical representation of the fitted model using the realtionship of *Equation 7* with the weights on each connection and bias has been illustrated in *Fig. 7*.

Predicted values of the dependent variable (l_1 of gauged site) have been provided in *Table 10*. *Table 10* also includes the estimated values of l_1 using QR analysis. The results show that ANN algorithm closely estimates l_1 for the gauged sites (Kalpani Deheri, Bara Tarnab, Khuderzai, Kalpani Saidabad, Jundi Tangi, Swat Ningolai and Bara Kohat Road) using available site characteristics as independent variables. ANN and QR model over estimate the values of l_1 for the sites (Wazir Ghari, Chinkar and Lund Khwar East) having smaller magnitude of AMPF. QR model under estimates l_1 for the sites (Jundi Utmanzai, Kalpani Saidabad and Jundi Tangi) having largre value of L-CV. The site Bara Tarnab have the larger magnitude of AMPF and its l-skewness is greater than L-CV QR model also under estimates its value. QR model provide reliable estimate of the values of l_1 for the sites (Kalpani Deheri, Khuderzai, Swat Ningolai and Bara Kohat Road). Similar results have been reported by (Dawson et al., 2006) showing the under and over estimation of floods at few gauged and ungauged sites for 10, 20 and 30 years return periods through multiple linear regression analysis and ANN. They proposed that this problem may be related to the accuracy and availability of the relevant site characteristics.
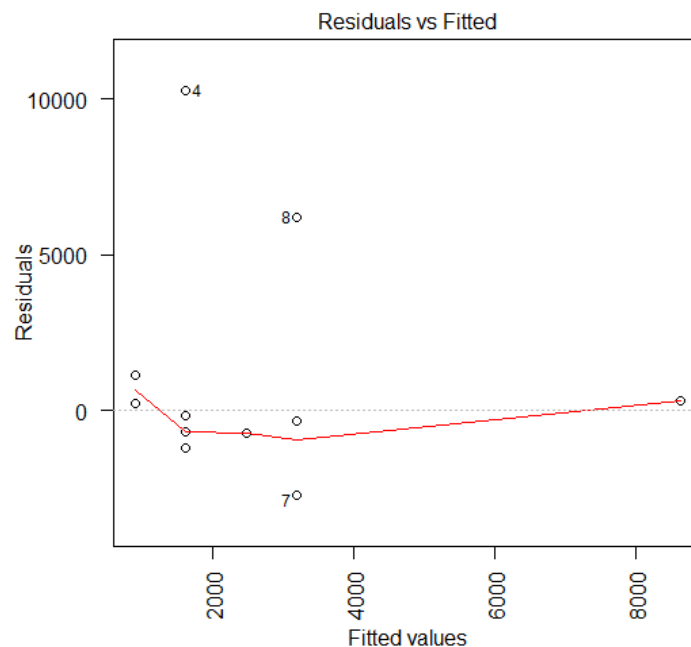

Figure 5. *Plot of fitted values of dependent variable, (first sample l-moment of annual maximum peak flows (l_1)) and residuals of the fitted regression model in Equation 6*
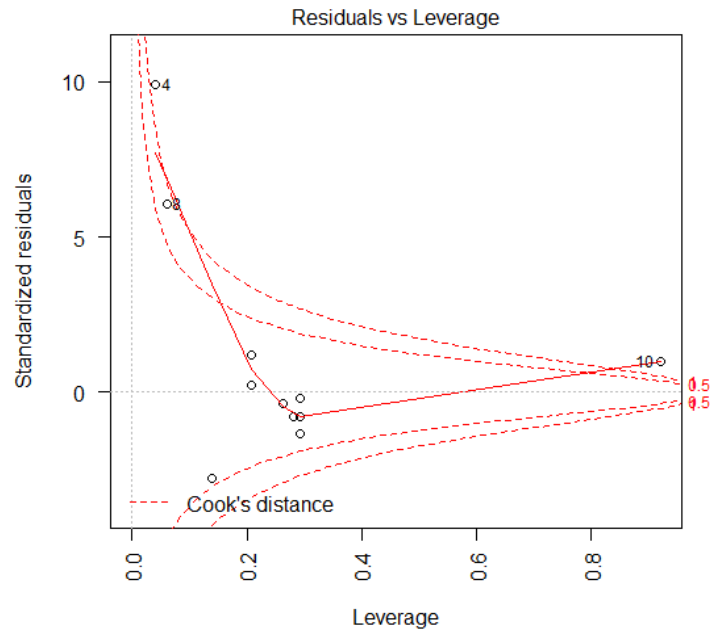
Khan et al.: A comparison of quadratic regression and artificial neural networks for the estimation of quantiles at ungauged sites in
regional frequency analysis
- 6952 -

Figure 6. *Diagnostic plot of fitted regression model (Equation 6)*
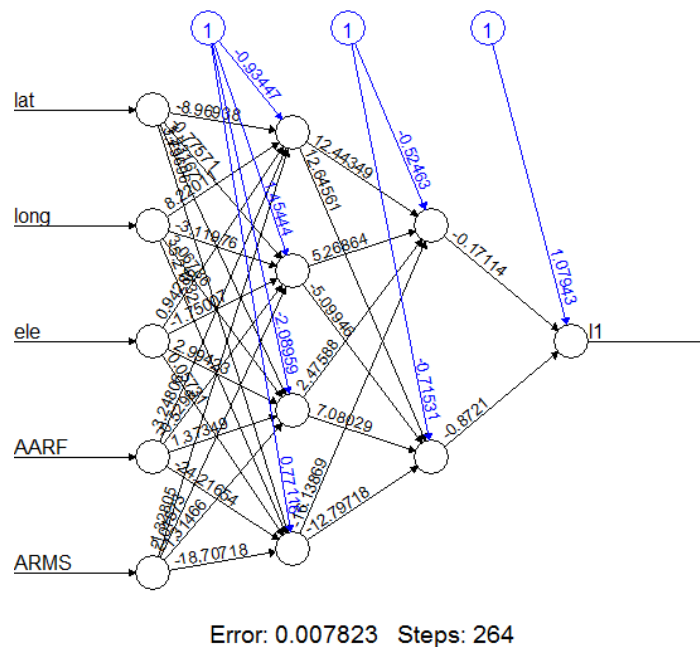

Error: 0.007823   Steps: 264

Figure 7. *Graphical representation of artificial neural networks (ANN) procedure adopted for the area under study to show the convergence of the model*

The performance of the QR model is not affective for the gauge site having smaller and larger magnitude of AMPF and for those sites also that show very high variation in the data set. As compared to QR model ANN estimates are more reliable and accurate accept the sites have smaller magnitude of AMPF. In *Fig. 8*, fitted values of dependent variable have been plotted against the observed values of the dependent variable. All the values predicted through ANN lie on the straight line relative to the predicted values

Khan et al.: A comparison of quadratic regression and artificial neural networks for the estimation of quantiles at ungauged sites in
regional frequency analysis
- 6953 -

through QR. Therefore, the observed and the predicted values of the dependent variable through ANN are quite close to each other, especially, for the sites having larger magnitude of AMPF.

Table 10. *Estimated values of l_1 of gauged sites through artificial neural networks (ANN) and quadratic regression (QR)*

| S.No.	Site name	l_1 (observed)	l_1 (fitted) using QR	l_1 (fitted) using ANN
1	Kalpani Deheri	2856.61	2854.886	2697.19
2	Wazir Ghari	426.47	1220.572	845.78
3	Chinkar	922.27	1220.572	1306.46
4	Bara Tarnab	11884.18	1220.572	11651.88
5	Khuderzai	1758.37	2110.622	1880.46
6	Jundi Utmanzai	2052.57	486.3447	842.53
7	Lund Khwar East	484.08	2854.886	841.43
8	Kalpani Saidabad	9408.82	2854.886	9587.09
9	Jundi Tangi	1104.65	486.3447	1199.7
10	Swat Ningolai	8933.68	8473.219	8874.14
11	Bara Kohat Road	1453	1220.572	1461.24
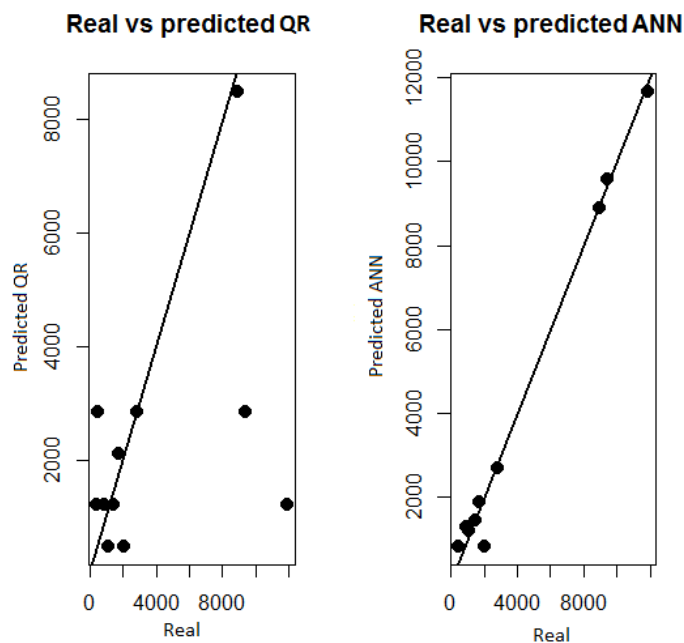
Figure 8. *Comparison of fitted values of site statistic l_1 (first sample l-moment of annual maximum peak flows) through artificial neural networks (ANN) and quadratic regression (QR)*

Assessment analysis

The two estimation methods (ANN and QR) have been used for T-year flood quantiles estimation at ungauged sites within the homogeneous region. The development of these two methods is theoretically justified in the study, still an assessment analysis is requisite to check the relative accuracy of estimates of the T-year flood quantiles for gauged, poorly

Khan et al.: A comparison of quadratic regression and artificial neural networks for the estimation of quantiles at ungauged sites in
regional frequency analysis
- 6954 -

gauged and ungauged sites. To do so, RMSE, MAPE and MAE have been calculated using the information of *Table 10* (observed and fitted values of each site of the homogeneous region) and the results are given in *Table 11*. Their corresponding formulas are:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{l}_i - l_i)^2}{n}} \qquad (Eq.8)$$

$$MAPE = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{|l_i - \hat{l}_i|}{|l_i|}\right) \times 100 \qquad (Eq.9)$$

$$MAE = \frac{\sum_{i=1}^{n}|l_i - \hat{l}_i|}{n} \qquad (Eq.10)$$

The results of *Table 11* show that the estimates through ANN are more accurate than QR as values of RMSE, MAPE and MAE calculated by using *Equation 8*, *Equation 9* and *Equation 10* respectively are relatively low. Therefore ANN can be preferred over QR to predict the flood quantiles for ungauged and poorly gauged sites of the study region.

Table 11. *Assessment measures for validation of the estimates through Artificial Neural Networks (ANN) and Quadratic Regression (QR)*

	RMSE	MAPE	MAE
ANN	432.12	27.17%	293.25
QR	3887.42	94.65%	2173.83

Note: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE)

Cross validation of estimated quantiles

The estimates of quantiles for various return periods through RFFA for gauged and ungauged sites are statistically sound but their practical validation is still requisite. A comparison of quantiles estimates obtained using index flood method, QR model and ANN for 10, 20 and 100 year return periods with highest values of observed/historic AMPF (first, second and third as per their order of magnitude along with their year of occurrence) at various sites, has been illustrated in *Table 12*. Due to high variation in the observed AMPF of each site three highest values are selected for the historic comparison of predicted quantiles.

A flood of 2010 has been considered as one of the worst in Pakistan's history (Government of Pakistan, 2016). The current study region is one of the affected areas of this heavy flood. There are four sites Chinkar, Jundi Utmanzai, Kalpani Saidabad and Bara Kohat Road where the highest values of the series of AMPF has been observed in 2010 (as shown in *Table 12*). For these sites, the predicted quantiles for 100 year return period using index flood mehtod for gauged sites and through ANN for ungauged sites are quite close to the highest obserseved AMPF for 2010. At site Jundi Tangi, second highest value of AMPF has been observed in 2010 and similar magnitude of flood has been predicted for 100 year return period through index flood and ANN. For the site Lund Khwar East, third highest value of AMPF has been observed in 2010 and same magnitude

of flood has been predicated for 20 year return period through index flood and 10 year return period for ANN. Similar trends of the observed and predicted values are obvious from the results of *Table 12*. This shows the accuracy and reliability in the predictive ability of the adopted procedure.

Table 12. *Cross validation of the estimated flood quantiles through index flood method, Quadratic Regression (QR) model and Artificial Neural Networks (ANN)*

S.No	Site name	Highest values of observed AMPF (year)			Estimated quantiles through index flood			Estimated quantiles through QR			Estimated quantiles through ANN		
		1st Highest	2nd Highest	3rd Highest	10	20	100	10	20	100	10	20	100
1	Kalpani Deheri	10845 (1995)	9380 (1997)	7915(1988)	6239	10106	26575	6236	10100	26559	5891	9542	25092
2	Wazir Ghari	3080 (1998)	2784 (1982)	849 (1994)	931	1508	3967	2666	4318	11355	1847	2992	7868
3	Chinkar	7500 (2010)	5625 (2011)	1387(2008)	2014	3262	8579	2666	4318	11355	2854	4622	12154
4	Bara Tarnab	159100 (2013)	20300(2005)	16912(1991)	25958	42043	110559	2666	4318	11355	25451	41222	108399
5	Khuderzai	13952 (1984)	9000 (1997)	5000 (1985)	3840	6220	16358	4610	7467	19635	4107	6653	17494
6	Jundi Utmanzai	19433 (2010)	11360(2008)	6465 (2011)	4483	7261	19095	1062	1720	4524	1840	2981	7838
7	Lund Khwar East	1891 (1997)	1735 (2006)	1546 (2010)	1057	1712	4503	6235	10100	26559	1838	2977	7828
8	Kalpani Saidabad	52912 (2010)	44321(2013)	39259(2015)	20551	33286	87531	6235	10100	26559	20941	33917	89190
9	Jundi Tangi	18679 (2008)	8382 (2010)	5837 (2007)	2412	3908	10276	1062	1720	4524	2621	4244	11161
10	Swat Ningolai	52098 (2016)	44870(2013)	28403(2014)	19513	31605	83110	18508	29976	78827	19384	31395	82557
11	Bara Kohat Road	11698 (2010)	5688 (1991)	5120 (1997)	3173	5140	13517	2666	4318	11355	3192	5170	13594

The results of *Table 12* show that the estimates, obtained using index flood method, are quite close to the historical values of AMPF at all most all of the sites. The results, obtained using QR model (for ungauged sites), have also been compared with the highest observed values of gauged sites. The comparison reveals that the estimates are comparable for the sites Kalpani Deheri, Wazir Ghari, Chinkar, Khuderzai, Swat Ningolai and Bara Kohat Road for smaller return periods with in the span of data. Significant deviations in the estimates have been observed for the sites Bara Tarnab, Jundi Utmanzai, Lund Khwar East, Kalpani Saidabad and Jundi Tangi for the smaller and longer return periods. The performance of QR estimates for the sites having large magnitude of observed AMPF has been observed as less efficient (sites Bara Tarnab and Kalpani Saidabad).

Estimates obtained through ANN are closed to the highest values of observed/historic AMPF of each site. Moreover, for the sites having large magnitude of observed AMPF, the results of ANN have been observed very close to the highest historical values of AMPF for smaller as well as longer return periods. Based on the historical comparison, ANN can be considered as more reliable relative to QR method to estimate quantiles for ungauged sites for smaller as well as larger return periods.

Summary, findings and conclusions

For a developing state with agro centered economy and facing problems of water shortage along with high variations in the stream flows, the importance of precise flood estimates becomes vast, especially for the small streams and rivers which originates with in the country. Some of the conclusions and key findings of the study are given below:

i. Few critical assumptions associated to the observed data series at various sites for RFFA has been tested and validated through various statistical tests. The results have revealed that the observed data at each site is random, independent, homogeneous and free of regular trends. Moreover, the results of the discordancy measure have shown that none of the site is discordant in the set of eleven gauging sites.

ii. Descriptive statistics have shown that there exist high variations, skewness and kurtosis in the observed AMPF of eleven gauging sites. However, the values of L-kurtosis are relatively smaller than the L-skewness. One possible reason of these variations is the irregular patterns of the monsoon rainfall, as flood flows in Pakistan depend mostly on the heavy monsoon rainfall. Similar findings have been reported by (Hussain, 2017) for the sites of major rivers of Punjab, Pakistan.

iii. The geographical locations of eleven sites reveal that the sites are geographically contiguous; therefore, it may be possible that they belong to the similar cluster with respect to their site characteristics. Formal heterogeneity measure, based on 1000 simulations assuming four parameters kappa distribution as regional probability distribution, confirms that the cluster of eleven sites is definitely homogeneous. GPA distribution has been identified as suitable regional distribution for the region under study. Regional quantiles have been estimated using quantile function of the GPA distribution. Estimated flood quantiles for 10 years return period or more at all sites are greater than the average value of observed AMPF of the respective sites.

iv. For flood quantiles estimation at ungauged sites, QR with robust estimation method and ANN have been used to develop a functional relationship between l_1 (mean of observed AMPF) of each site and their respective available site characteristics. The results show that the estimates using ANN are accurate and reliable as compared to QR analysis especially for longer return periods.

v. To assess the practical validity of the provided estimates, for various return periods, through L-moments based RFFA using index flood method, ANN and QR, a comparison has been illustrated using first, second and third highest values of the recorded AMPF at eleven gauging sites. The estimates of RFFA are closer to the recorded highest values of AMPF at all sites for different return periods. Moreover, the estimates obtained through ANN (though comparison has been provided for gauged sites only) are also very close to the highest historic values of AMPF of each corresponding site for different return periods. Therefore, ANN would be a preferred approach as compared to QR for flood quantiles estimation at ungauged sites (especially for longer return periods) within the homogeneous region.

The study contributes in terms of promoting non-linear relationship of the most influential site characteristic with the observed AMPF with robust estimation method emphasizing the complete justifications related to the estimation procedure of the regression modeling. Moreover, the introduction of ANN to obtain flood quantiles for

ungauged sites in Pakistan and comparison with the historical record to investigate the practical validity of the resultant estimates. The results of this study not only provide assistance to the officials dealing with flood risks management but will also be useful for management of agriculture water and design capacity of existing and proposed hydrologic structures in the study region. For instance, a proposed project of the provincial government of KPK is the site Bara Dam (an ungauged site of the study area).

REFERENCES

[1]   Abrahart, R., Kneale, P. E. See LM (eds.) (2004): Neural networks for hydrological modeling. – CRC Press.

[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): 2345-2353.

[3]   Ahmad, I., Fawad, M., Akbar, M., Abbas, A., Zafar, H. (2016): Regional Frequency Analysis of Annual Peak Flows in Pakistan Using Linear Combination of Order Statistics. – Polish Journal of Environmental Studies 25(6): 2255-2264.

[4]   Alam, J., Muzzammil, M., Khan, M. K. (2016): Regional flood frequency analysis: comparison of L-moment and conventional approaches for an Indian catchment. – ISH Journal of Hydraulic Engineering 22(3): 247-253.

[5]   Anilan, T., Satilmis, U., Kankal, M., Yuksek, O. (2016): Application of Artificial Neural Networks and regression analysis to L-moments based regional frequency analysis in the Eastern Black Sea Basin, Turkey. – KSCE Journal of Civil Engineering 20(5): 2082-2092.

[6]   Aydoğan, D., Kankal, M., Önsoy, H. (2016): Regional flood frequency analysis for Çoruh Basin of Turkey with L-moments approach. – Journal of Flood Risk Management 9(1): 69-86.

[7]   Aziz, K., Rahman, A., Fang, G., Shrestha, S. (2014): Application of artificial neural networks in regional flood frequency analysis: a case study for Australia. – Stochastic environmental research and risk assessment 28(3): 541-554.

[8]   Batool, Z. (2017): Flood Frequency Analysis of Stream Flow in Pakistan Using L-Moments and TL-Moments. – International Journal of Advance Research, Ideas and Innovations in Technology 3(4): 136-142.

[9]   Bradley, J. V. (1968): Distribution-free statistical tests. – No. 04; QA278. 8, B7.

[10]   Dawson, C. W., Abrahart, R. J., Shamseldin, A. Y., Wilby, R. L. (2006): Flood estimation at ungauged sites using artificial neural networks. – Journal of hydrology 319(1-4): 391-409.

[11]   Development Advocate Pakistan (2017): Water Security in Pakistan: Issues and Challenges 3(4).   http://www.pk.undp.org/content/pakistan/en/home/library /hiv_aids/development-advocate-pakistan--volume-3--issue-4.html.

[12]   Glaves, R., Waylen, P. R. (1997): Regional flood frequency analysis in southern Ontario using L-moments. – The Canadian Geographer/Le Géographe canadien 41(2): 178-193.

[13]   Government of Pakistan (2016): Annual flood report 2016. – Ministry of Water and Power, Office of the Chief Engineer Advisor and Chairman, Federal Flood Commission, Islamabad

Khan et al.: A comparison of quadratic regression and artificial neural networks for the estimation of quantiles at ungauged sites in
regional frequency analysis
- 6958 -

http://www.ffc.gov.pk/download/AFR/Annual%20Flood%20Report%202016.pdf.
(Accessed June 2018).

[14] Griffis, V. W., Stedinger, J. R.(2007): The use of GLS regression in regional hydrologic analyses. – Journal of Hydrology 344(1-2): 82-95.

[15] Hailegeorgis, T. T., Alfredsen, K. (2017): Regional flood frequency analysis and prediction in ungauged basins including estimation of major uncertainties for mid-Norway. – Journal of Hydrology: Regional Studies 9: 104-126.

[16] Hashmi, H. N., Siddiqui, Q. T. M., Ghumman, A. R., Kamal, M. A. (2012): A critical analysis of 2010 floods in Pakistan. – African Journal of Agricultural Research 7(7): 1054-1067.

[17] Hirsch, R. M., Helsel, D. R., Cohn, T. A., Gilroy, E. J. (1992): Statistical analysis of hydrologic data. – In: Maidment, D. R. (ed.) Handbook of Hydrology, Chapter 17. McGraw-Hill, New York.

[18] Hjelmfelt, A. T., Wang, M. (1996): Predicting runoff using artificial neural networks. – In Proceedings of the International Conference on Hydrology and Water Resources, New Delhi, India, December 1993, Springer, Dordrecht: 233-244.

[19] Hosking, J. R. M., Wallis, J. R. (1997): Regional frequency analysis: an approach based on L-moments. – Cambridge University Press.

[20] Hussain, Z., Pasha G. R. (2009): Regional flood frequency analysis of the seven sites of Punjab, Pakistan, using L-moments. – Water resources management 23(10): 1917-1933.

[21] Hussain, Z. (2011): Application of the regional flood frequency analysis to the upper and lower basins of the Indus River, Pakistan. – Water resources management 25(11): 2797-2822.

[22] Hussain, Z. (2017): Estimation of flood quantiles at gauged and ungauged sites of the four major rivers of Punjab, Pakistan. – Natural hazards 86(1): 107-123.

[23] Hussain, Z., Shahzad, M. N., Abbas, K. (2017): Application of regional rainfall frequency analysis on seven sites of Sindh, Pakistan. – KSCE Journal of Civil Engineering 21(5) 1812-1819.

[24] Jingyi, Z., Hall, M. J. (2004): Regional flood frequency analysis for the Gan-Ming River basin in China. – Journal of hydrology 296(1): 98-117.

[25] Khan, S. A., Hussain, I., Hussain, T., Faisal, M., Muhammad, Y. S., Mohamd Shoukry, A. (2017): Regional Frequency Analysis of Extremes Precipitation Using L-Moments and Partial L-Moments. – Advances in Meteorology.

[26] Kumar, R., Chatterjee, C., Panigrihy, N., Patwary, B. C., Singh, R. D. (2003a): Development of regional flood formulae using L-moments for gauged and ungauged catchments of North Brahmaputra river system. – Journal of the Institution of Engineers India Civil Engineering Division 84: 57-63.

[27] Kumar, R., Chatterjee, C., Kumar, S., Lohani, A. K., Singh, R. D. (2003b): Development of regional flood frequency relationships using L-moments for Middle Ganga Plains Subzone 1 (f) of India. – Water Resources Management 17(4): 243-257.

[28] Kumar, R., Chatterjee, C. (2006): Closure to "Regional Flood Frequency Analysis Using L-Moments for North Brahmaputra Region of India". – Journal of Hydrologic Engineering 11(4): 380-382.

[29] Landi, A., Piaggi, P., Laurino, M., Menicucci, D. (2010): Artificial neural networks for nonlinear regression and classification. – In Intelligent Systems Design and Applications (ISDA), 10th International Conference IEEE: 115-120.

[30] Liu, D., Yuan, Y., Liao, S. (2009): Artificial neural network vs. nonlinear regression for gold content estimation in pyrometallurgy. – Expert Systems with Applications 36(7): 10397-10400.

[31] Maier, H. R., Dandy, G. C. (2000): Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. – Environmental modelling & software 15(1): 101-124.

[32] Malekinezhad, H., Zare-Garizi, A. (2014): Regional frequency analysis of daily rainfall extremes using L-moments approach. – Atmósfera 27(4): 411-427.

[33] Modarres, R. (2008): Regional frequency distribution type of low flow in North of Iran by L-moments. – Water Resources Management 22(7): 823-841.

[34] Muhammad, F., Afreen, S. (2007): Modeling annual maximum peak flows at various dams and barrages in Pakistan. – Journal of Hydrology and Hydromechanics 55(1): 43-53.

[35] National Electric Power Regulatory Authority (NEPRA) (2018): Hydel Potential in Pakistan. – http://www.nepra.org.pk/Policies/Hydel%20Potential%20in%20Pakistan.pdf.

[36] Noto, L. V., La Loggia, G. (2009): Use of L-moments approach for regional flood frequency analysis in Sicily, Italy. – Water resources management 23(11): 2207-2229.

[37] Ouali, D., Chebana, F., Ouarda, T. B. (2017): Fully nonlinear statistical and machine-learning approaches for hydrological frequency estimation at ungauged sites. – Journal of Advances in Modeling Earth Systems 9(2): 1292-1306.

[38] Pakistan Bureau of Statistics (2018). http://www.pbs.gov.pk/content/agriculture-statistics.

[39] Pakistan Meteorological Department (2012): The implementation of diagnostic study for 2010 flood and extreme moon soon rains 2011 in Pakistan under sustainable development through peace building, governance and economic recovery in KP and support landslide IDPs in Hunza Nagar and Gilgit district when UNDP surves as implementing partner. – http://www.pmd.gov.pk/reports/flood_diagnostic_2010_2011.pdf. (Accessed June 2018).

[40] Pokhrel, J. (2002): Regional flood frequency analysis for the island of Newfoundland, Canada using L-Moments. – Doctoral dissertation, Memorial University of Newfoundland.

[41] Requena, A. I., Ouarda T. B., Chebana, F. (2017): Flood Frequency Analysis at Ungauged Sites Based on Regionally Estimated Stream flows. – Journal of Hydrometeorology 18(9): 2521-2539.

[42] Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1985): Learning internal representations by error propagation (No. ICS-8506). – California Univ San Diego La Jolla Inst for Cognitive Science.

[43] Saf, B. (2009): Regional flood frequency analysis using L-moments for the West Mediterranean region of Turkey. – Water Resources Management 23(3): 531-551.

[44] Shahzadi, A., Akhter, A. S., Saf, B. (2013): Regional frequency analysis of annual maximum rainfall in monsoon region of Pakistan using L-moments. – Pakistan Journal of Statistics and Operation Research 9(1): 111-136.

[45] Shamseldin, A. Y., Nasr, A. E., O'Connor, K. M. (2002): Comparison of different forms of the multi-layer feed-forward neural network method used for river flow forecast combination. – Hydrology and earth system sciences 6(4): 671-684.

[46] Shu, C., Ouarda, T. B. M. J. (2007): Flood frequency analysis at ungauged sites using artificial neural networks in canonical correlation analysis physiographic space. – Water Resources Research 43(7).

[47] Sivakumar, B., Singh, V. P. (2012): Hydrologic system complexity and nonlinear dynamic concepts for a catchment classification framework. – Hydrology and Earth System Sciences 16(11): 4119-4131.

[48] Wald, A., Wolfowitz, J. (1943): An exact test for randomness in the non-parametric case based on serial correlation. – The Annals of Mathematical Statistics 14(4): 378-388.

[49] Yang, T., Xu, C. Y., Shao, Q. X., Chen, X. (2010): Regional flood frequency and spatial patterns analysis in the Pearl River Delta region using L-moments approach. – Stochastic Environmental Research and Risk Assessment 24(2): 165-182.