Instantaneous Flood Events Data Analyses in Northern Pakistan

Kashif Hussain, M.Sc.
National Development consultants, Commercial Broadway, Lahore
Muhammad Shahab, B.Sc.
Department of Civil Engineering,
University of Engineering & Technology, Peshawar, Pakistan

Abstract

In large watersheds of Pakistan, a proper flood intensity assessment is a most important concern for many water management applications. The present study investigates the possible best-fit probability distributions technique of uttermost flood events using more than 30 years data (except Mastuj Bridge station) from 15 stream flow stations in northern Pakistan, operated and maintained by Surface Water Hydrology Project (SWHP), WAPDA. Different statistical distribution model, were used to estimate design floods for risk planning and the design of important civil infrastructure. Method of moments was used for estimation of parameters of selective distributions. Four goodness-of-fit statistics test K, A2, $\chi 2$ and Root mean square error (RMSE) were applied to evaluate the most appropriate probability distribution. P3, LP3, GEV, and LN were determined best fit distribution model. Among the eight-probability distribution's P3 fitted 33.3% of the stations, LP3 also fitted 33.3% of the stations, GEV fitted 26.7% of the stations and LN fitted 6.7% of the stations. The more practical result for peak flood were calculated for recurrence intervals of 5-year, 25-year, 50-year, 100-year, 500-year and 1000-years respectively. In future before planning and designing any project such as Hydropower, irrigation systems, Dike, flood disaster assessment and management at Indus and adjacent rivers, P3, LP3, GEV and LN distributions results should be considered more reliable on these locations.

Keywords: Flood frequency analysis, Indus basin, Instantaneous flood peak, Goodness of fit (GOF) test, return period

Introduction

Water related issues in Pakistan are most crucial challenges for the policymakers, hydrologist and water resources engineers in the country

(Akhtar, et al., 2008) and extreme flood events are one of the primary natural causes of economic losses in Asia. Natural hazard or calamity and all kind of flood are probably the most frequent and devastating, causing serious loss damages of economic, livelihood of people, and biodiversity (Mohapatra & Singh, 2003; Wang et al., 2011; Sharifi et al., 2012 ; Bandyopadhyay et al., 2016). The probabilities of occurrence of extreme events are determined by frequency analysis method. The great importance of flood frequency analysis (FFA) is to colligate the magnitude of uttermost events to their frequency of occurrence by using different probability distributions technique (Alam et al., 2018) Climate patterns and flood magnitudes vary from basin to basin as well as from selective stream gauging station to station. However, watershed homogeneity can also be found. Taking annual extreme floods as the extreme event, Faisal et al., (2015) found that in northern Punjab the Gumbel's distribution is giving best fit. Amin et al., (2016) studied 6 climate station annual maximum rainfall complete that study area before found that the normal and log-Pearson type 3 (LP3) model was the best-fit distribution. (U. N. Ahmad, 2011), found generalized logistic was the best distribution (GLO) for L moment in Negeri Sembilan, Malaysia for annual maximum stream flows over stations. (Hussain, 2011) found Pearson Type 3 distribution best for the Kalabagh site and generalized logistic was the best distribution (GLO) for the Kotri at Indus River. (Mazhar Saeed, Ahmad, & Nabi, 2012) workout the flood frequency for Jhelum Basin's homogeneous regions and found Gumbel was the best distribution based on Chi-square test. (Bandyopadhyay et al., 2016) were performed vulnerability assessment in the Ajay River, Eastern India and workout LPT-3 best distribution for flood modeling. The magnitude of the flood was estimated by most affectively distribution Generalized Pareto, and Log Pearson type 3 of different period (Singo, Kundu, Odiyo, Mathivha, & Nkuna, 2012); (Bezak, Brilly, & Šraj, 2014); (Khattak et al., 2016); (Kamal et al., 2016) and, (Benameur, Benkhaled, Meraghni, Chebana, & Necir, 2017). The method of moments, method of L-Moments or maximum likelihood estimators can be used estimation parameters of flood frequency distributions (Bezak et al., 2014). The parameters of limited range can be calculated by good technique method of moment (MOM), while L-Moments are most vast and unbiased used (Weglarczyk, Strupczewski, & Singh, 2002). Pakistan has an agriculture-based economy where the role of water (precipitation, snow and glaciers) is important and Indus basin system is a major source of livelihoods of more than 138 million people which depend on irrigated agriculture in the country. The events of supper floods such as in 2015 glaciers lake outburst flood (GLOF) in Chitral River, catastrophic flood of 2010 in Swat, Indus and downstream of Kunhar Rivers and most supper flood of 1992 in Kunhar River of Jehlum Basin were affected large number of agriculture land, ecosystems, biodiversity and livelihood of people. It is

important to extract the patterns of tremendous flood events to evaluate the flood risk factors which can be used to build long-term measures to save economic damages and lives. Statistical analyses of the available instantaneous peak flood (IPF) data are needed for proposing effective control or mitigation of flood measures for downstream area. It also tries to work out the best technique in fitting instantaneous peak flood (IPF) data. The instantaneous peak flood (IPF) data could be verified using suitable distribution technique as well as the goodness of fit (GOF) tests (Zhang, 2002). The objective of this paper focuses on two essential issues. First to evaluate the best-fit models for instantaneous peak flood events frequency analysis in northern Pakistan. Second to calculate the expected flood peak over return period 5-year, 25-year, 50-year, 100-year, 500-year and 1000-years.

Study Area and Available Data

In the present study five river basins have been selected to determine the flood magnitude of mountainous area of northern Pakistan. These watersheds include Upper Indus Catchment (UIC), Chitral basin, Golen Gol basin, Swat Basin and Kunhar Basin. The Indus River is a longest transboundary river in Asia and it fed the world's largest conterminous irrigation system. It originates at Singi Kahad near Mansarwar Lake on Kailas Parbat in the frozen deserted Tibetan Plateau (Ali, 2013). Gilgit River is a right bank tributary of Indus River and confluence to Indus just 8.7 km upstream of Jaglot Town. Hunza River is a left bank tributary of Gilgit basin and entered the Gilgit River some 40 kilometers above the Gilgit–Indus joint. One of its left bank tributary Astore River drained from Nanga Parbat eastern face and joined the Indus just 11 km downstream of the Bunji Town. The Indus River is about 2115 km long up to Tunsa Barrage in Pakistan. Swat and Chitral River is the largest left bank tributary of transboundary Kubal River Basin (KRB). Swat River originates from the high mountains of Swat Kohistan. The total catchment area of the Swat River is 14690 Km^2 and total length of the main Swat basin reach is 274 km, which accounts for approximately 22.3% of the total drainage area of river Kabul. Chitral River which is also known as Mastuj River originates from Chikzar, Chiantar and Zindi Kharam Glaciers in Hindu Kush mountains. Most of the upper watershed remains covered with snow and glaciers in winter season. Kunhar River is one of the largest tributaries of Jhelum River basin located in the northern side of Pakistan and it originates from the Lulusar Lake. Approximately 65% to the total discharge of the Kunhar River added from snowmelt (Mahmood, Jia, & Babel, 2016). The selected rivers basin as demarcated was shown in **Error! Reference source not found.**. The waters of the selected rivers basin and their tributaries were also having great importance from the perspective of water resources management, flood risk analysis, hydrological monitoring and for power

generation, as 29 % electricity contributed from hydropower into the National Power Grid of Pakistan (Tarar, Ahmad, Ahmad, & Majid, 2018).
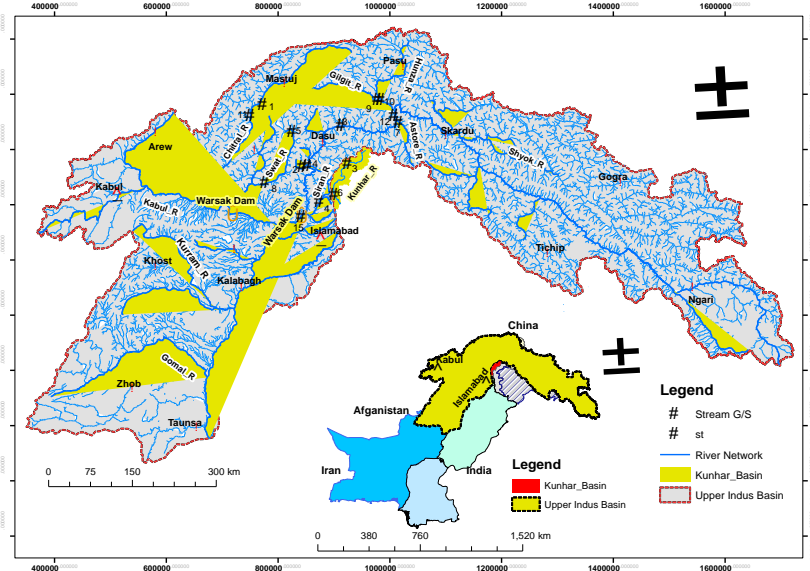

Figure 4. *Study area map*

In the present study, the hydrological peak data consisted of the instantaneous flood at 15 gauging stations.

Time series of instantaneous peak records were collected by the Surface Water Hydrology Project (SWHP) of (WAPDA), using current meters for discharges measurement of 15 different locations in northern area of the country. The data was provided as the instantaneous peak in meter cube per second (m^3/s). This analysis uses more than 30 years of data; however, there is one station (Mastuj Bridge) which did not have 30 years of available data (***Error! Reference source not found.***).

Table 1. *Stream gauging sites and flood records used in the flood frequency analyses*

Sr. #	Station Name	River Name	Data Years	Catchment Area	Elevation
				km^2	masl
1	Mastuj Bridge	Golen Gol	23	520	1649
2	Karora	Khan Khwar	32	572	910
3	Naran/Kaghan	Kunhar	57	1107	2035
4	Phulra	Siran	40	1154	732
5	Kalam	Swat	48	2020	1921
6	Talhata	Kunhar	55	2354	826
7	Doyian	Astore	35	4040	1583
8	Chakdara	Swat	47	5776	676
9	Gilgit	Gilgt	49	12095	1430
10	Dainyor	Hunza	43	13157	1350
11	Chitral	Chitral	52	11396	1500
12	Bunji	Indus	44	142900	1188

13	Shatail	Indus	32	154670	1040
14	Besham	Indus	47	163178	580
15	Tarbela	Indus	48	169600	458

Methodology

Indus and its tributaries and adjacent catchments in Pakistan are most important for hydropower and water resources projects. An enormous hydropower, water resources and other infrastructure proposed on catchments selected in this study, some of which completed in early 1980s and many of them under constructions or ready to constructions. So that, a proper method was selected to evaluate the best-fit probability distribution for a certain gauging sites for flood frequency analysis. The most commonly used distribution method in instantaneous flood peak (IFP) analyses are described in this section.

Method of Moments (MOM)

The method of moments (MOM) is a more natural technique to estimate the parameter of a common probability distribution function. It is a statistical technique for parameters estimators and based sample moments to match with the corresponding distribution moments (Vivekanandan, 2015). To simulation the parameters by using MOM; the sample mean X of a random variable, σ standard deviation of a random variable and γ coefficient of skewness of a random variable are determined by the following equations (Alam et al., 2018):

$$\bar{\mu} = \frac{1}{n}\sum_{j=1}^{n} Xj \tag{1}$$

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(Xj - \mu)^2} \tag{2}$$

$$\gamma = \frac{n\sum_{j=1}^{n}(Xj - \mu)^3}{(n-1)(n-2)S^3} \tag{3}$$

Where $\bar{X}$ is the mean of the observed flood, n is the total number of observed data (j), j is the observed data number and X_j is the observed data, while σ is a standard deviation of flood data and γ coefficient of skewness of the observed flood data.

Normal

For flood mitigation or control measures a normal distribution model is better because of economic losses which directly associated with design

floods (Rao & Hamed, 2000). The probability density function f(x) and cumulative distribution function F(x) for a normally distributed random variable x is written by,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2} (x - \mu)^2\right] \qquad (4)$$

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\exp\left[-\frac{1}{2\sigma^2} (x - \mu)^2\right]\right) dx \qquad (5)$$

Where μ and σ are the parameters of the normal distribution model and x random variable can take any value of range $-\infty < x < \infty$. The maximum discharge of desired return period in normal distribution model can calculated using;

$$X_T = \mu + u\sigma \qquad (6)$$

Where X_T is the computed discharge of desired return period, μ is the mean, σ is the standard deviation and u is the standard normal variate replaced with K_T, which depends on the return period and probability distribution. The value of u can be calculated by equation below (Abramowitz et al., 1965)

$$u = W - \frac{2.515517 + 0.802853w + 0.0110328W^2}{1 + 1.432788W + 0.189269W^2 + 0.001308W^3} + \in (P) \qquad (7)$$

Log-Normal

The probability density and cumulative functions of two parameter log-normal distributed variable x is given by:

$$f(x) = \frac{1}{x\sigma_Y\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_Y^2} (\ln(x) - \mu_Y)^2\right] \qquad (8)$$

$$F(x) = \frac{1}{\sigma_Y\sqrt{2\pi}} \int_0^x \left(\frac{1}{x}\exp\left[-\frac{1}{2\sigma_Y^2} (\ln(x) - \mu_Y)^2\right]\right) dx \qquad (9)$$

Where σ_Y and μ_Y is the standard deviation and mean for the LN2 distribution model. The flood peak against desired return period can be computed by using LN2 is given by:

$$X_T = e^{\mu + u\sigma} \qquad (10)$$

Pearson Type 3

The probability density distribution and cumulative distribution function of P3 are expressed as:

$$f(x) = \frac{1}{|\alpha|\Gamma(\beta)}[\left(\frac{x-\xi}{\alpha}\right)^{\beta-1}]exp[-\frac{(x-\xi)}{\alpha}] \qquad (11)$$

$$F(x) = \frac{1}{|\alpha|\Gamma(\beta)}\int_{\xi}^{x}[\left(\frac{x-\xi}{\alpha}\right)^{\beta-1}]\exp\left[-\frac{(x-\xi)}{\alpha}\right]dx \qquad (12)$$

The range of x value in P3 distribution can be take $\Upsilon<x<\infty$, whereas α scale parameter and it is generally positive or negative, therefore negative value of α can make distribution upper bounded and not suitable for any hydrological event analyses. Where β shape and ξ location parameters. The flood peak of different return period can be calculated with equation below:

$$X_T = \alpha\beta + \gamma + K_T\sqrt{\alpha^2\beta} \qquad (13)$$

Where K_T in above equation is the frequency factor corresponding to a desired return period.

Log-Pearson Type 3
. LP3 probability density and cumulative distribution functions (Griffis & Stedinger, 2007):

$$f(x) = \frac{1}{|\alpha|x\Gamma(\beta)}[\left(\frac{\ln(x)-\xi}{\alpha}\right)^{\beta-1}]exp[-\frac{(\ln(x)-\xi)}{\alpha}] \qquad (14)$$

$$F(x) = \frac{1}{|\alpha|\Gamma(\beta)}\int_{0}^{x}\frac{1}{x}[\left(\frac{\ln(x)-\xi}{\alpha}\right)^{\beta-1}]\exp\left[-\frac{(\ln(x)-\xi)}{\alpha}\right]dx \qquad (15)$$

The flood peak using distribution model against any selected return period (T) can be calculated using

$$X_T = \exp^{(\mu + K_T\sigma)} \qquad (16)$$

Two Parameter Gamma Distribution
It is a very special case of the Pearson type 3 distribution, the two-parameter gamma probability density function can be expressed as:

$$f(x) = \frac{1}{\alpha^\beta\Gamma(\beta)}x^{\beta-1}e^{-(x/\alpha)} \qquad (17)$$

The cumulative distribution function for two-parameter gamma model can be express as:

$$F(x) = \frac{1}{\alpha^\beta\Gamma(\beta)}\int_{0}^{x}x^{\beta-1}e^{-(x/\alpha)}dx \qquad (18)$$

The range of random value x are $0<x<\infty$, where α and β are scale and location parameter of two-parameter gamma distribution.

Gumbel

Gumbel distribution or extreme value type 1 is one of the special cases of GEV distribution has been extensively reported in literature of statistics hydrology, is often used in statistical method predicting magnitude of worst hydrological events such as floods. For Hydrologic flood data analysis GUM distribution were found best-fit model in Finland and Spain (Salinas, Castellarin, Kohnová, & Kjeldsen, 2014) and (Mamman, Martins, Ibrahim, & Shaba, 2017) workout the extreme value type 1 best-fit distribution for Kainji Reservoir. Extreme value type 1 probability density and cumulative distribution functions are given by:

$$f(x) = \left(\frac{1}{\alpha}\right) \exp\left[-\left(\frac{x-\beta}{\alpha}\right) - \exp^{-\left(\frac{x-\beta}{\alpha}\right)}\right] \tag{19}$$

$$F(x) = \exp\left[-\exp^{-\left(\frac{x-\beta}{\alpha}\right)}\right] \tag{20}$$

The value range of variable x can take $-\infty < x < \infty$ and (α, β) are the scale and location parameters for the GUM distribution. The Gumbel distribution can be used to simulate the flood peak against desired return period using given equation.

$$X_T = \beta - \alpha \ln\left[-\ln\left(1 - \frac{1}{T}\right)\right] \tag{21}$$

Generalize Extreme Value

For flood frequency analysis GEV distribution is a well-known standard method in the UK (Selaman, Said, & Putuhena, 2007) and it is best fit for flood data recommended in many European country including Spain, Italy, Austria and Germany (Salinas et al., 2014). The probability density and cumulative functions of GEV distribution are given by:

$$f(x) = \frac{1}{\alpha}[1 - k\left(\frac{x-\beta}{\alpha}\right)]^{\frac{1}{k}-1} \exp\left\{-[1 - k\left(\frac{x-\beta}{\alpha}\right)]^{\frac{1}{k}}\right\} \tag{22}$$

$$F(x) = \exp\left\{-[1 - k\left(\frac{x-\beta}{\alpha}\right)]^{\frac{1}{k}}\right\} \tag{23}$$

Where α and β are scale and location parameter of GEV distribution and k shape parameter which depends only coefficient of skewness.

Weibull

The application of Weibull distribution in hydrology and meteorology have been reported widely and commonly used distribution for analysis of risk, and reliability system of life time as well as for flood frequency (Stephenson, 1998). For low extreme flows analysis used of Weibull was recommend extensively. The probability density and cumulative functions of Weibull distribution are expressed as:

$$f(x) = \left(\frac{\alpha}{\beta}\right)\left(\frac{x}{\beta}\right)^{\alpha-1}\exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right] \tag{24}$$

$$F(x) = 1 - \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right] \tag{25}$$

Where α and β are shape parameter and scale parameter for the W2 distribution, with the reported range $x > 0$; $\alpha > 0$, $\beta > 0$.

Goodness-of-Fit Test

A few goodness-of-fit tests such as Kolmogorov-Smirnov, Anderson-Darling and chi-square have been conducted to check the best distribution model. A Q-Q plots as graphical test and root mean square test were also applied to determine the best fit model of each location. Therefore, the Cramer Von Mises, Kolmogorov-Smirnov (K-S), and the Anderson-Darling (A-D) are most used EDF tests (Arshad et. al, 2003; Seier, 2002).

Kolmogorov-Smirnov (K-S) Test

In nature K-S is a nonparametric test widely used in software packages especially statistical and important in real world to measures the extensive contradictoriness between the observed sample value and hypothesized distribution. A random sample x_1, x_2, x_3 …...xi are, suppose from some distribution with sustained distribution function $F(x)$ and empirical cdf can expressed as:

$$F_n(x) = \frac{I \le x}{n} \tag{26}$$

Where n is representing the size of the random sample and I is the number observation less than or equal to x (Evans, Drew, & Leemis, 2017). D_n is a statistic of K–S test, which define largest vertical distance between $F_n(x)$ and $F(x)$ for all x values can be computed by

$$D_n^+ = \max_{i=1,2,..n}\left\{\frac{i}{n} - F(x_i)\right\} \tag{27}$$

$$D_n^- = \max_{i=1,2,..n}\left\{F(x_i) - \frac{i-1}{n}\right\} \tag{28}$$

Where xi is the ith order statistic,

$$D_n = \max\{D_n^+, D_n^-\} \tag{29}$$

An algorithm also provided by (Drew, Glen, & Leemis, 2000) for computing the cdf of Dn statistic in case all parameters of hypothetical cdf $F(x)$ are known.

Anderson-Darling (A-D) Test

The most well-known and powerful statistics tests of the EDF is AD tests (Shin, Jung, Jeong, & Heo, 2012) in contrast with the Kolmogorov-Smirnov test which have comparatively less power and can give better

discrimination between distributions (M. I. Ahmad & Sinclair, 1988). It was first introduce by Anderson–Darling (T.W & D. A, 1954) to minimize errors in the stochastic progress and have extreme importance in frequency analysis of hydrology and meteorological data. The mathematical process can be written as follows (Zeng, Wang, & Wu, 2015): Sorting the samples X (x_1, x_2… x_n) by ascending order, and storing it to a new vector X′ (x'_1, x'_2… x'_n), and for computational purposes, the A-D statistics (A_n^2) following equation form can be used:

$$A_n^2 = -n - \frac{1}{n} \sum_{i=1}^{n} [(2i-1)\log F(x'_i) + (2n+1-2i)\log F(x'_i)] \quad i = 1,2,\dots,n \tag{30}$$

Chi-Square Test

The χ^2 GoF test is a most commonly used formal and practical test (Huang, Lee, & Ting, 2008) has vast applicability in statistical work to evaluate the best-fit distribution model for hydrologic data analysis. The Chi-square test appears quite different to suppose that the observations of sample is large sufficient so that the chi-square distribution give a better estimation as the distribution of test statistic (Ghosh, Roy, & BISWAS, 2013). The Chi-Squared statistic is given by:

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{31}$$

Where O_i is the observed number of cases in category i, and E_i is the expected number of cases in category i.

Root Mean Square Error (RMSE)

Root mean square error were used as 4^{rth} best-fit distribution model in this study. Finally, the best-fit distributions model for observed IFP at different stations of Indus basin as well as at adjacent catchments are workout using this approach. A RMSE can be computed for data set as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^{n} (x_i - X)^2} \tag{32}$$

Where xi denotes the model predicted value and X denotes the observed sample value. In many cases of distribution model sensitivity studies only RMSE were used, a detailed elucidation is not critical because varieties of a similar model will have comparative error distributions. When assessing distinctive models using a unique metric, contrasts in the error distributions

become progressively essential (Chai & Draxler, 2014). The RMSE gives a generally large weight to extensive error by squaring them.

Graphical Test

A Graphical test is also an effective technique to establish best-fit distribution model for flood data. Here Q-Q plots are means of comparing observed value to estimated value. The Q-Q plot can also guide us how well the assessed distribution is performing over the scope of the quintiles. The non-exceedance probability $P_{i:n}$ plotting position can compute of each $x_{i:n}$ from different plotting position formulas, which were used for the particular distributions. Plotting position formula used here have the following form:

$$P_{i:n} = \frac{i - 0.4}{n + 0.2} \tag{33}$$

Where xi observed value are ranked in increasing order, and denoted from $x_{1:n}$ to $x_{n:n}$, and where n is the total number of observed values. To construct the Q-Q plot, each $x_{i:n}$ is paired with $y_{i:n}$, which is simulated from the supposed cumulative distribution function (CDF), F(x). "The set of points ($x_{i:n}$, $y_{i:n}$) is plotted on normal graph with a 1:1 straight line extending from the origin. Theoretically, all points should fall on the 1:1 line if the assumed CDF is the true distribution" (Bourque, Nguyen, & Tao, 2002).

3.3 Return Period

To compute the return period or recurrence interval is one of the important goals of IFP frequency analysis. The floods of large magnitude would not occur at any fixed pattern and naturally have large recurrence interval. A selective flood value x with a recurrence interval (T) may be exceeded once in T year. The probability of exceedance in given year is:

$$P(X_T > x) = \frac{1}{T} \tag{34}$$

The cumulative probability of non-exceedance F (XT) is given by:

$$F(X_T) = P(X_T \le x) = 1 - P(X_T > x) = 1 - \frac{1}{T} \tag{35}$$

The return period for the design of critical hydraulic structure varies as a hydraulic structure importance in study area or downstream of the study area, that is for desired goal of flood mitigation or control, flood risk management/damage reduction, touristic, and socio-economic development. Sometime it is very crucial to oversize the hydraulic structure to reduce the damage risk in case of some worst flood events (Urías & Garcia, 2007).

Results and Discussion

The study area Indus Basin, adjacent rivers and catchments is surrounded by steep mountainous range with elevation peak varied between 400 to 8500 m in the Himalayan, Karakoram and Hindu ranges. The snow and glacier melt from these steep mountains, combined with massive monsoon rainfall, contribute flows to these rivers. The main scope of this paper to evaluate the best-fit probability distribution for all stream gauging station which yield the instantaneous flood peak (IFP) for return periods of 5, 25, 50, 100, 500 and 1000 years. These estimated flood peak can provide useful information and guidance for flood disaster management, civil infrastructure designing, policy making and decisions purposes. The computed return period of supper and catastrophic flood events would be more helpful and can be used to evaluate the damage risk level.

Selecting the Best-Fit Distribution

The data analyzed in our study showed the flood peak range in all stream gauging stations varied widely due to basin geometric and climate characteristics. The upper Indus River basin and selected surrounding catchments in northern area of Pakistan has the highest amount of flow originated from snow and glacier melting and rainfall, whereas in few gauging station precipitations also contribute missive runoff especially during monsoon season e.g., Tarbela and Talhata stations as recorded a high instantaneous flood peak (IFP) 23648 and 2274 m^3/s respectively over the past more than 30 years. Q-Q plots were created of selected stations to examine the distributions with fit degree on right tail but in many cases, it is very difficult to work out the best fit distribution from Q-Q plot's visual display. Q-Q Plots for Chitral and Besham Qila gauging station are illustrated as an example in **Error! Reference source not found.** and **Error! Reference source not found.**. The major concern of fitting the probability distribution for all sites is to represent the low probability most properly, worst event on the extreme right tail. ''By using a Q-Q plot the level of fit on the extreme right tail can be studied'' (Alam et al., 2018). On other hand, there are enormous test technique and estimation method reported in many statistical books and research studies to check the best distributions for flood analysis. Root mean square error (RMSE) one of them, were used to check the fit distribution from observed and estimated random number generated at 15 stations data. The best probability distribution was selected based on minimum statistics error computed with RMSE method for the individual data set of probability distributions. K-A, A-D and Chi-square tests were also used to check the best probability distribution of study areas. Table 2 shows the selected gauging stations names, results of statistical summary and the best-fit distribution results with ranks using RMSE, K-S, A-D, and Chi-Square test. The smallest

251

rank from all four goodness-of-fit tests were chooses to best fit probability distribution which are worked out for each station. The best-fit distribution model based on test statistics were shows in last column of Table *2*. The both graphical observation and numerical test were found most important to check or select the best-fit probability.

Sr. #	Station Name	Mean (X)	SD (σ)	C_V	Coeff. of Skew (γ)	Best-Fit Test Statistic Result								Best Fit Model
						K-S Test		A-D Test		Chi-Test		RMSE		
						Dist.	Statistics	Dist.	Statistics	Dist.	Statistics	Dist.	Statistics	
1	Mastuj Bridge	64	48	0.74	2.244	LP3	0.125	LP3	0.2633	P3	0.1166	LP3	10.61	LP3
2	Karora	423	427	1.01	2.656	GEV	0.0697	GEV	0.2172	LP3	0.3672	P3	47.52	GEV
3	Naran/Kaghan	274	146	0.53	3.041	GEV	0.0722	P3	0.1811	LP3	1.4268	GEV	25.82	GEV
4	Phulra	674	308	0.46	0.631	GEV	0.0889	LP3	0.1898	LP3	0.9174	GAM	36.12	LP3
5	Kalam	432	199	0.46	4.662	LN	0.1098	P3	0.8063	GEV	1.7532	LP3	87.15	LN
6	Talhata	616	343	0.56	2.79	P3	0.0621	P3	0.2207	LP3	1.5576	GEV	51.44	P3
7	Doyian	603	211	0.35	0.509	P3	0.0585	P3	0.1602	Gum	0.2108	GEV	23.21	P3
8	Chakdara	1072	864	0.81	5.158	GEV	0.2212	LP3	6.8353	LP3	5.228	LP3	388.29	LP3
9	Gilgit	1297	369	0.28	2.58	P3	0.0722	P3	0.4229	LN	2.4787	GEV	66.48	P3
10	Dainyor	1688	369	0.22	0.7	P3	0.1131	LP3	0.4854	LN	2.4746	LP3	66.4	LP3
11	Chitral	1186	271	0.23	1.742	P3	0.0437	P3	0.083	LP3	0.4351	GEV	16.08	P3
12	Bunji	7998	1110	0.14	0.556	LP3	0.0779	P3	0.3301	GEV	3.5958	LP3	177.35	LP3
13	Shatail	9414	1958	0.21	1.879	GEV	0.0818	P3	0.2583	P3	0.1314	GEV	303.83	GEV
14	Besham	10997	2247	0.2	2.408	P3	0.0732	P3	0.3785	GEV	3.1954	LP3	568.52	P3
15	Tarbela	11564	2548	0.22	2.331	P3	0.0666	GEV	0.2895	GEV	3.7742	LP3	699.11	GEV

Table 2. *Statistical and best-fit results of the selected stations*
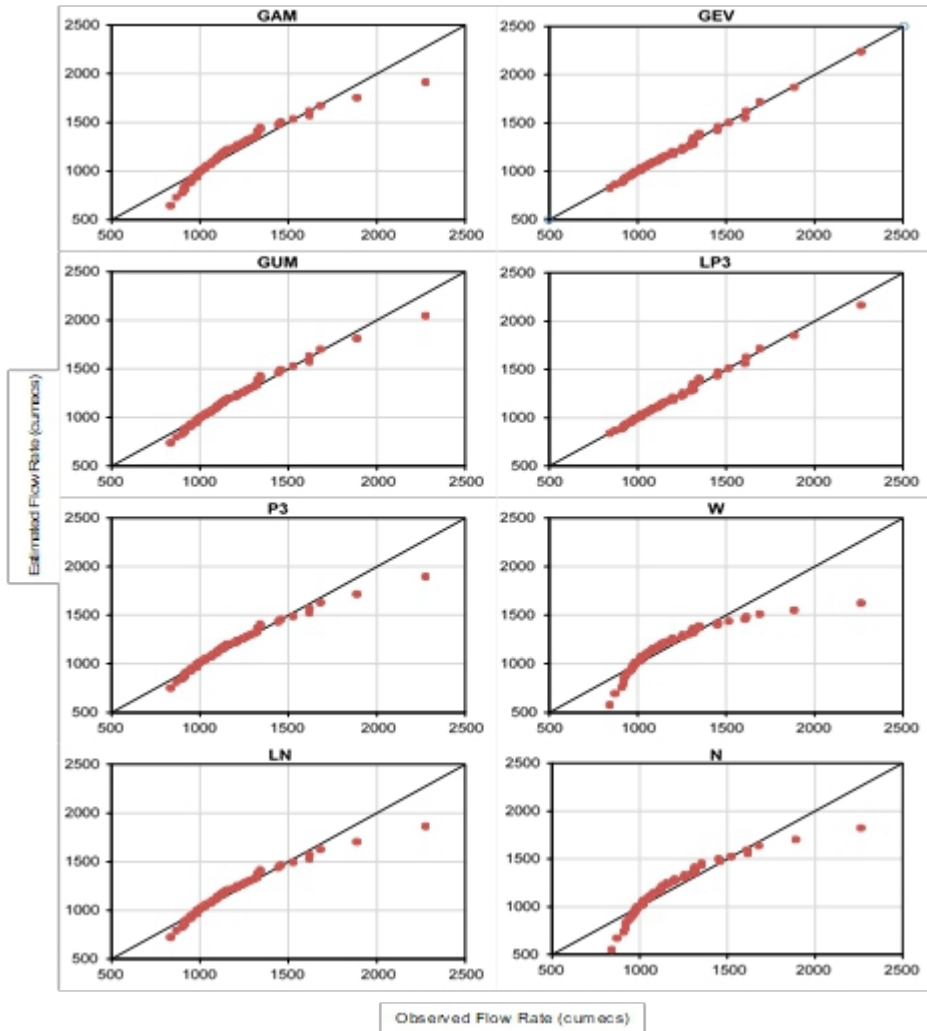
252

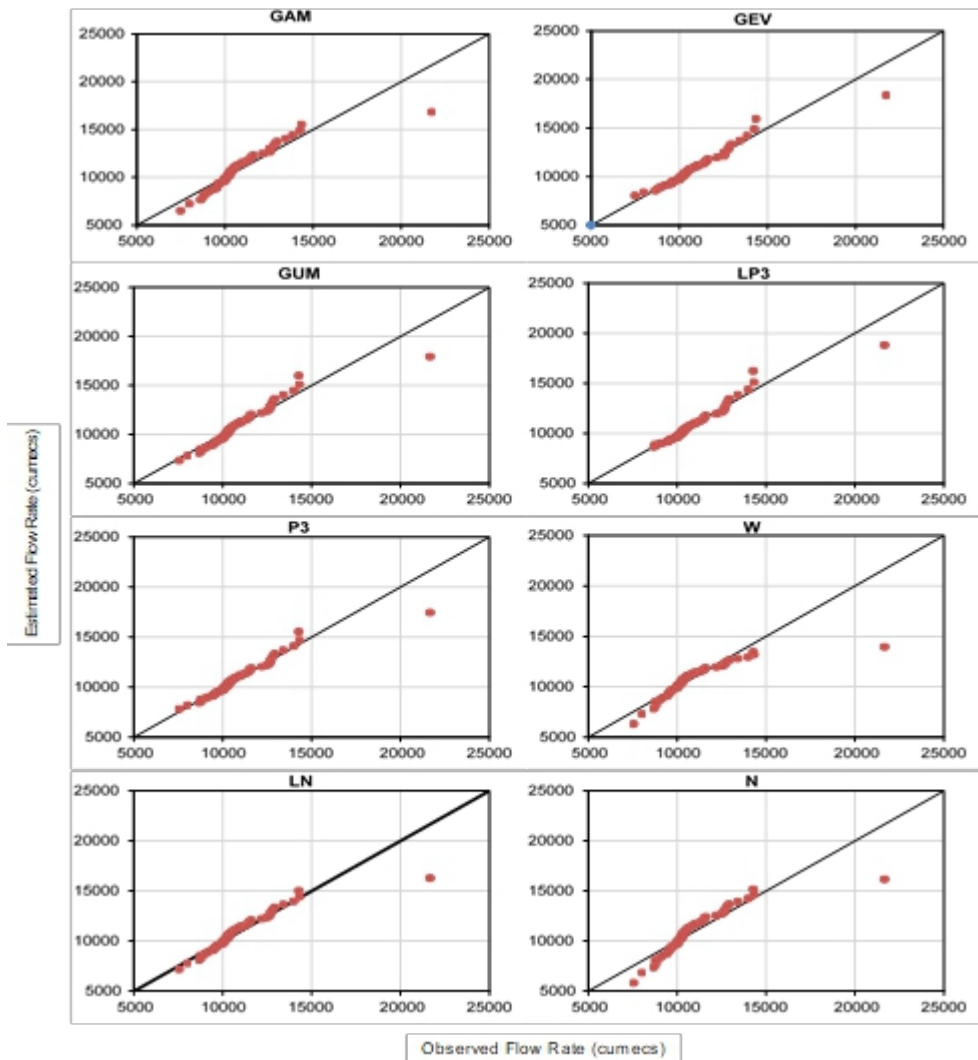**Figure 2.** *Q-Q plot of all distributions at Station: Chitral*

Figure 3. Q-Q plot of all distributions at Station: Besham

The computed statistical parameters such as mean, standard deviation (SD), coefficient of variation (CV), skewness and kurtosis for the Instantaneous flood peak (IFP) data of 15 stations are summarized in Table 2. The distribution for the Instantaneous flood peak (IFP) was found to be positively skewed at all different stations. The Instantaneous flood peak data set indicates that Chakdara station flood peak was strongly positively skewed. The coefficient of skewness of 20% stations are less than 1 thus can be regarded as low skew and 80% station lie within the range 1 to 5.15 thus are highly skewed. The most fit probability distributions statistics of four test D, A^2, χ^2 and RMSE were calculated for each station data set reported into Table 2. P3, LP3 and GEV obtained first rank according to Kolmogorov Smirnov,

Anderson Darling, Chi-Square and RMSR tests at different stations. It was observed that at 33% stations P3 probability distribution, at 33% station LP3 probability distribution, at 26.7% stations GEV probability distribution and at 6.7 LN distribution found best. Among all distribution P3 and LP3 yielded the most cases of best-fit distributions, while second and third amount of best fits distributions was found GEV and LN respectively. The LN and Gamma distributions gave some best fit results among few stations. However, statistics results of Gum, N and W based on four tests were not found best for all stations.

Return Period Results

The worst floods occurred due to massive monsoon precipitation in the Indus River and lower Kunhar basin has posed emphasis for flood frequency analysis using best fit distribution model. The instantaneous flood peak frequency of 5-year, 10-year, 25-year, 50-year, 100-year, 500-year, and 1000 –years return period carried out of all gauging station are illustrated in Figure 5. All study stations from 1 to 15 presented in **Error! Reference source not found.** were taken as horizontal axis and floods magnitude of different return period represents in vertical axis. The estimated flood magnitude of return period 5-year, 10-year, 25-year, 50-year, 100-year, 500-year, and 1000-years are shown from top to bottom vertically. P3, LP-3 and GEV distributions appears to be the most appropriate fit among eight probability distributions, flood magnitude estimated by best-fitted distribution P3 at Besham, station number 14 of return period 5, 25, 50, 100, 500 and 1000 year are 12199.9 m^3/s, 16067.7 m^3/s, 17796 m^3/s, 19545.6 m^3/s, 23664.1 m^3/s and 25455 m^3/s respectively.
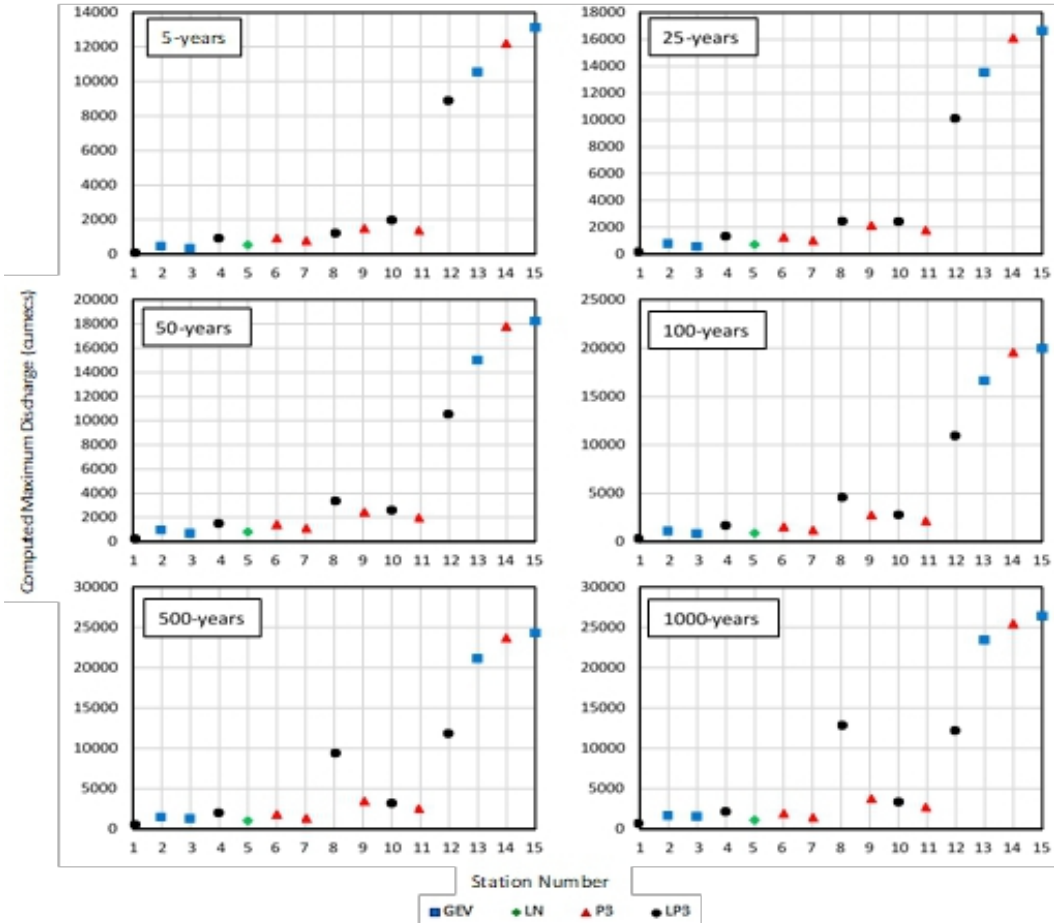
Figure 5*. Return period results of all station*

Conclusion

The instantaneous flood peak patterns varied across all stations due to geographical location, river and catchment geometric characteristics and surrounding climate are most important in the variation of the instantaneous flood peak pattern of the selected basins. Monsoon season contributed massive runoff in Indus River at Tarbela, Siran River at Phulra and Kunhar River at Talhatta. In term of determining the best-fit distribution to model the instantaneous flood peak for a certain station of selected basins, the result varied significantly. Method of moment were used to estimation the parameters of the distribution and the graphical procedure, the Q-Q plot, K, A^2, χ^2 and RMSE tests were used to estimation of best fit probability distribution. To civil structure designing and long-term risks analysis and management return period is most important were also computed using best fit distributions model. In many cases all four test statistics yielded the more than one best fit distribution type. Thus, based on low statistics error the best

fit distribution for each station were determine and ranked. The best fit distribution of each station was selected with lowest statistics error results as well as lowest rank scores. Among eight distributions model P3, LP3 and GEV found best fit distributions of different stations. According to minimum statistics error results, P3 fitted 33.3% of the stations, LP3 also fitted 33.3% of the stations, GEV fitted 26.7% of the stations and LN was fitted 6.7% of the stations. Flood assessment, management, monitoring and mitigation or control is crucial for civil infrastructures safety as well as environmental sustainability, particularly within river downstream area. Return period calculation is an effort in current study for all locations to provide an appropriate planning tool for civil structure designer and administrators. Here, 5-year, 25-year, 50-year, 100-year, 500-year and 1000-year return periods of instantaneous flood peak (IFP) return levels were illustrated. In future before planning and designing any project such as Hydropower, bridge, irrigation systems, Dike, flood disaster assessment and management at Indus and adjacent rivers, P3, LP3 and GEV distributions results should be considered more reliable on these locations.

References:
1. Abramowitz, M., Stegun, I. A., & Miller, D. (1965). Handbook of Mathematical Functions With Formulas, Graphs and Mathematical Tables (National Bureau of Standards Applied Mathematics Series No. 55). Journal of Applied Mechanics, Vol. 32, p. 239. https://doi.org/10.1115/1.3625776
2. Ahmad, M. I., Sinclair, C. D., Spurr, B.D. (1988). Assessment of Flood Frequency Models Using Empirical Distribution Function Statistics. Water Resources Research. 24(8), 1323–1328.
3. Ahmad, U. N. (2011). Flood Frequency Analysis of Annual Maximum Stream Flows using L-Moments and TL-Moments Approach. Applied Mathematical Sciences, 5(5), 243–253.
4. Akhtar, M., Ahmad, N., & Booij, M. J. (2008). The impact of climate change on the water resources of Hindukush-Karakorum-Himalaya region under different glacier coverage scenarios. Journal of Hydrology, 355(1–4), 148–163. https://doi.org/10.1016/j.jhydrol.2008.03.015
5. Alam, M., Emura, K., Farnham, C., & Yuan, J. (2018). Best-Fit Probability Distributions and Return Periods for Maximum Monthly Rainfall in Bangladesh. Climate, 6(1), 9. https://doi.org/10.3390/cli6010009
6. Ali, A. (2013). Indus Basin Flood Mechanisms, Impacts, and Management. https://doi.org/10.1016/j.earscirev.2007.07.002
7. Amin, M. T., Rizwan, M., & Alazba, A. A. (2016). A Best-Fit

Probability Distribution for the Estimation of Rainfall in Northern Regions of Pakistan. Open Life Sciences, 11(1), 432–440. https://doi.org/10.1515/biol-2016-0057

8. Arshad, M.; Rasool, M.T.; Ahmad, M.I. (2003). Anderson-Darling and Modified Anderson-DarlingTest for Generalized Pareto Distribution. Pakistan Journal of Applied Sciences, 3(2) 85-88.

9. Bandyopadhyay, S., Ghosh, P. K., Jana, N. C., & Sinha, S. (2016). Probability of flooding and vulnerability assessment in the Ajay River, Eastern India: implications for mitigation. Environmental Earth Sciences, 75(7). https://doi.org/10.1007/s12665-016-5297-y

10. Benameur, S., Benkhaled, A., Meraghni, D., Chebana, F., & Necir, A. (2017). Complete flood frequency analysis in Abiod watershed, Biskra (Algeria). Natural Hazards, 86(2), 519–534. https://doi.org/10.1007/s11069-016-2703-4

11. Bezak, N., Brilly, M., & Šraj, M. (2014). Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. Hydrological Sciences Journal, 59(5), 959–977. https://doi.org/10.1080/02626667.2013.831174

12. Bourque, A., Nguyen, V.T.V., & Tao, D. (2002). On Selection of Probability Distributions for Representing Annual Extreme Rainfall Series. Global Solutions for Urban Drainage, (January 2002), 1–10. https://doi.org/doi:10.1061/40644(2002)250

13. Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. Geoscientific Model Development, 7(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

14. Drew, J. H., Glen, A. G., & Leemis, L. M. (2000). Computing the cumulative distribution function of the Kolmogorov-Smirnov statistic. Computational Statistics and Data Analysis, 34(1), 1–15. https://doi.org/10.1016/S0167-9473(99)00069-9

15. Evans, D. L., Drew, J. H., & Leemis, L. M. (2017). The distribution of the Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling Test Statistics for Exponential Populations with Estimated Parameters. International Series in Operations Research and Management Science, 247, 165–190. https://doi.org/10.1007/978-3-319-43317-2_13

16. Faisal, M., Nabi, G., Abbas, T., Yaseen, M., Boota, M. W., & Lahore, T. (2015). An Appraisal of Flood Occurrence and Varability in Streamflows of Nullah Deg, North East Punjab, Pakistan. 27(5), 4369–4372.

17. Ghosh, S., Roy, M. K., & Biswas, S. C. (2013). Determination of the Best Fit Probability Distribution for Annual Extreme Precipitation in Bangladesh. European Journal of Scientific Research, 103(JUNE),

391–404. https://doi.org/10.5923/j.ajms.20160604.05

18. Griffis, V. W., & Stedinger, J. R. (2007). Log-Pearson Type 3 Distribution and Its Application in Flood Frequency Analysis. I: Distribution Characteristics. Journal of Hydrologic Engineering, 12(5), 482–491. https://doi.org/10.1061/(asce)1084-0699(2007)12:5(482)

19. Huang, Y. P., Lee, C. H., & Ting, C. S. (2008). Improved estimation of hydrologic data using the chi-square goodness-of-fit test. Journal of the Chinese Institute of Engineers, 31(3), 515–521. https://doi.org/10.1080/02533839.2008.9671406

20. 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. https://doi.org/10.1007/s11269-011-9839-5

21. Kamal, V., Mukherjee, S., Singh, P., Sen, R., Vishwakarma, C. A., Sajadi, P., Rena, V. (2016). Flood frequency analysis of Ganga river at Haridwar and Garhmukteshwar. Applied Water Science, 7(4), 1979–1986. https://doi.org/10.1007/s13201-016-0378-3

22. Khattak, M. S., Anwar, F., Saeed, T. U., Sharif, M., Sheraz, K., & Ahmed, A. (2016). Floodplain Mapping Using HEC-RAS and ArcGIS: A Case Study of Kabul River. Arabian Journal for Science and Engineering, 41(4), 1375–1390. https://doi.org/10.1007/s13369-015-1915-3

23. Mahmood, R., Jia, S., & Babel, M. S. (2016). Potential impacts of climate change on water resources in the kunhar river basin, Pakistan. Water (Switzerland), 8(1), 1–24. https://doi.org/10.3390/w8010023

24. Mamman, M. J., Martins, O. Y., Ibrahim, J., & Shaba, M. I. (2017). Evaluation of best-fit probability distribution models for the prediction of inflows of Kainji Reservoir, Niger State, Nigeria. Air, Soil and Water Research, 10. https://doi.org/10.1016/j.ecocom.2017.11.003

25. Mazhar Saeed, M., Ahmad, I., & Nabi, G. (2012). Flood frequency analysis of homogeneous regions of Jhelum River Basin. International Journal of Water Resources and Environmental Engineering, 4(5), 144–149. https://doi.org/10.5897/IJWREE12.013

26. Mohapatra, P. K., & Singh, R. D. (2003). Flood management in India. Natural Hazards, 28(1), 131–143. https://doi.org/10.1023/A:1021178000374

27. Rao, A.R.; Hamed, K.H. (2000). Flood Frequency Analysis. CRC Press.

28. Salinas, J. L., Castellarin, A., Kohnová, S., & Kjeldsen, T. R. (2014). Regional parent flood frequency distributions in Europe – Part 2: Climate and scale controls. Hydrology and Earth System Sciences,

18(11), 4391–4401. https://doi.org/10.5194/hess-18-4391-2014

29. Seier, E. (2002). Comparison of Tests for Univariate Normality Tests for Normality. Interstat, (1), 1–17. https://doi.org/10.1080/00949659208811407

30. Selaman, O. S., Said, S., & Putuhena, F. J. (2007). Flood Frequency Analysis for Sarawak using Weibull,Gringorten and L-Moments Formula. Journal The Institution of Engineers,Malaysia, 68(1), 43–52.

31. Sharifi, F., Samadi, S. Z., & Wilson, C. A. M. E. (2012). Causes and consequences of recent floods in the Golestan catchments and Caspian Sea regions of Iran. Natural Hazards, 61(2), 533–550. https://doi.org/10.1007/s11069-011-9934-1

32. Shin, H., Jung, Y., Jeong, C., & Heo, J. H. (2012). Assessment of modified Anderson-Darling test statistics for the generalized extreme value and generalized logistic distributions. Stochastic Environmental Research and Risk Assessment, 26(1), 105–114. https://doi.org/10.1007/s00477-011-0463-y

33. Singo, L. R., Kundu, P. M., Odiyo, J. O., Mathivha, F. I., & Nkuna, T. R. (2012). Flood frequency analysis of annual maximum stream flows for Luvuvhu River Catchment, Limpopo Province, South Africa. 16th SANCIAHS Hydrology Symposium, (March 2016), 1–3.

34. Singh, V.P. (1998). Entropy-Based Parameter Estimation in Hydrology. Water Science and Technology Library, (30), 184

35. T.W, A., & D. A, D. (1954). A Test Of Goodness Of Fit. Journal of the American Statistical Association, 49(268), 765–769.

36. Tarar, Z. R., Ahmad, S. R., Ahmad, I., & Majid, Z. (2018). Detection of sediment trends using wavelet transforms in the Upper Indus River. Water (Switzerland), 10(7), 1–17. https://doi.org/10.3390/w10070918

37. Urías, H. Q., & Garcia, H. (2007). Determination of the Relationship Between Precipitation and Return Periods to Assess Flood Risks in the City of Juarez , Mexico.

38. Vivekanandan, N. (2015). Flood frequency analysis using method of moments and L-moments of probability distributions. Cogent Engineering, 2(1). https://doi.org/10.1080/23311916.2015.1018704

39. Wang, Y., Li, Z., Tang, Z., & Zeng, G. (2011). A GIS-Based Spatial Multi-Criteria Approach for Flood Risk Assessment in the Dongting Lake Region, Hunan, Central China. Water Resources Management, 25(13), 3465–3484. https://doi.org/10.1007/s11269-011-9866-2

40. Weglarczyk, S., Strupczewski, W. G., & Singh, V. P. (2002). A note on the applicability of log-Gumbel and log-logistic probability distributions in hydrological analyses: II. Assumed pdf. Hydrological Sciences Journal, 47(1), 123–137. https://doi.org/10.1080/02626660209492912

41. Zeng, X., Wang, D., & Wu, J. (2015). Evaluating the Three Methods of Goodness of Fit Test for Frequency Analysis. Journal of Risk Analysis and Crisis Response, 5(3), 178. https://doi.org/10.2991/jrarc.2015.5.3.5

42. Zhang, J. (2002). Powerful goodness-of-fit tests based on the likelihood ratio. Journal of the Royal Statistical Society. Series B: Statistical Methodology, 64(2), 281–294. https://doi.org/10.1111/1467-9868.00337