Flood Frequency Analysis of River Swat Using Easyfit Model & Statistical Approach

Qasim Sarfaraz[1], Muhammad Masood[2*], Abdul Sattar Shakir[3], M. Kaleem Sarwar[2], Noor M. Khan[3], Aftab H Azhar[4]

1. *Three High Head Hydro Power Projects, WAPDA, Pakistan.*
2. *Centre of Excellence in Water Resources Engineering, University of Engineering & Technology, GT-Road, 54890 Lahore, Pakistan.*
3. *Department of Civil Engineering, University of Engineering & Technology GT-Road, 54890, Lahore, Pakistan*
4. *Pakistan Water & Power Development Authority (WAPDA)-Mangla Dam Org. Pakistan-10200*

* **Corresponding Author:** Email: masood@cewre.edu.pk

Abstract

Pakistan has faced a tragic and massive flood in 2010. The climate change is considered a major factor for such a devastating and severe monsoon. The widespread precipitation generated very high runoff in Indus, Kabul, Swat Chenab and Jhelum rivers. Swat River is a major tributary of the Indus Basin River system, located in between the foothills of Hindukush mountain range also known for its snowcapped peaks. The contribution of snowmelt, average ground water and average rainfall in the basin is 65%, 19% and 16% respectively. Average annual rainfall in this region is more than 1000 mm. The dominant sources of rainfall are westerlies and monsoon which contribute 45% and 55% respectively. The sharp flow peaks are generated due to lack of surface storage capacity and non-absorption of runoff in the catchment area of River Swat. Therefore floods are common in this basin, which portrays threat both to infrastructure & humans. The unprecedented flooding in 2010 destroyed Munda & Amandara Headworks at River Swat due to much higher peak flood than the design discharge. High flow velocities and flow energy instigated erosion and also damaged infrastructure such as roads, houses and bridges. Due to undulated terrain and accessibility, there is deficiency of precipitation and discharge recording stations. Consequently measures to avert losses from such events cannot be taken well in time. In this study, flood frequency analysis of River Swat at Chakdara Station was carried out using Gumbel's Extreme Value (Type-1), Normal, Log Normal and Log Pearson Type III distributions against 2, 5, 10, 20, 25, 50, 100, 500, 1000, and 10000 year return periods to estimate the occurrence of such flood events. Significance tests, Anderson Darling, Kolmogrov Smirnov and Chi-Squared were applied in order to assess the most effective Probability Distribution for the study area. It was observed that statistical distributions helped considerably in estimation of floods at sites of homogeneous regions with less or no data. The results of the study, based on the applied significance tests i.e. Kolmogrov Smirnov, Anderson Darling and Chi-Square, exhibited that Gen Extreme Value [Type-1] ranked superior. Therefore it is the best-fit distribution among the other applied distributions for the Study area. Considering the changes in the behavior of streams and patterns of flooding, it is recommended to review the criteria and design limits for structures in urban & rural areas as well as of river training works.

Key Words: Flood Frequency Analysis, River Swat, Statistical Approaches, Easyfit

1. Introduction

Floods are Natural events, characterized by flow, velocity, temporal & spatial dynamics, water depth and matter fluxes. Pakistan is one of the South Asian countries, which have faced several catastrophic disasters in last 66 years. During the period of 1950 to 2011 about twenty-one major flood events have been experienced by Pakistan [1]. Heavy concentrated rainfall during monsoon season and snowmelt cause flood in Pakistan. Flood of 2010 was the second worst ever flood in sense of devastation. Riverine flooding was observed in Southern areas while flash floods were observed in Northern areas of Pakistan and Northwestern parts of India [2].

Rain fall of July 2010 was five times more than usual rainfall during July in Pakistan. Flash floods were resulted in Kabul and Swat rivers due to this unprecedented rainfall. Heavy rainfall on Karakorum and Hindukush ranges accelerated glacial and snow melt which caused extraordinary flood in river Indus at Tarbela. Second rainfall spell produced another flood wave during the same year. These two flood events originated the longest persistent flood in the history of Pakistan. The Patterns of global weather were under the effect of EI Nino a year before, intimidated monsoon rainfall in Pakistan. Transition of EI Nino to opposite phenomenon of La Nino contributed unusual rain in Northwestern part of Pakistan. Climate variability and land use land cover changes are the basic causes towards increase of vulnerability and frequency of occurrence of floods during several years in Pakistan, especially in Khyber Pakhtunkhwa (KPK) Province [3].

Akhtar [4] by utilizing collected cross-sectional data set from 812 households examined localized floods impact on livelihood of farmers in Pakistan. The study concluded that a rural livelihood was severely affected. Moreover, the affected households have reduced food security levels due to lowered cereal crop yields resulting less income. Akiyuki [5] studied the relationship b/w poverty and floods and concluded that the poverty can be aggravated by floods, as the poor people tend to live in flood-prone areas. The research was based on questionnaire survey conducted in Bago city Myanmar.

Flood Frequency Analysis (FFA) estimates probable occurrence of specified flood event, for this assessment a best fit frequency distribution is carefully chosen which on the basis of previous characteristics and magnitude estimates the probability of occurrence of flood. This requires the best fit frequency model and the availability of historical record [6]. Analysis of Stream flow data is more important to acquire flood's probability distribution before any estimation [7]. Validity of FFA results depends upon a hypothesis, independent and identical in distribution of the series [8]. Different studies employed many statistical distributions in order to quantify likelihood and intensity of floods but the findings remain area specific and none gained worldwide universal acceptance [9].

Probability distribution of annual peak discharge is very important in determination of statistics of flows of different magnitudes. Which are then applied for design, planning and management of water resource related projects. Log-Normal, Weibull, Normal, Log Pearson Type 3, Pearson Type 3, Gamma and Extreme value distributions are most commonly used probability distributions in stochastic hydrology [10][11] [12]. Pearson and Log Normal distributions are considered comparatively better for peak flows and rainfall, while extreme value distribution and Weibull distributions are best for extremes hydrological variables [13][14].The option of statistical model and procedure to be adopted for parameter estimation may introduce uncertainty in estimates. Generalized extreme value distribution in combination of max likelihood estimation method involves largest uncertainty as compared to logpearson (Type -3) [15].

The selection of most suitable distribution has always remained a challenging job. Various flood distributions were investigated by McMahon and Srikanthan [16]. They applied "Log-Pearson Type 3 Distribution and observed its applicability to Flood Frequency Analysis of Australian Streams in Australia. Vogel and Wilson [17] performed a very extensive study related to Probability Density Function of Mean, Maximum and Minimum annual stream flows in USA using flow observations of 1455 stations. Sasan [18] applied an artificial neural network (ANN) model for regional flood frequency analysis across New South Wales (in Australia) by utilizing stream flow data from 88 gauging stations. In this study, 3 to 8 predictor variables like drainage area, rainfall intensity, shape factor, river slope, annual rainfall, potential evapotranspiration, stream density, fraction forested area and drainage area were tested. The study concluded that an ANN model with higher number of predictor variables, doesn't always improve performance of regional flood frequency analysis models.

Khan [19] investigated various frequency distributions for Narmada River in India using monthly peak flow data using GEV type 1, lognormal, Normal and Log Pearson (type 3) distributions. He observed that Normal Distribution was found the most suitable for Garudeshwar Station using monthly maximum stream flow data. Wisam [20] conducted analysis on River ZAB using Gumbel, Log-Normal type3, Log Pearson type 3 and Pearson type 3 statistical models and concluded Log Normal type 3 as best fit for this river. Sathe [21] applied Log Pearson Type 3 on the catchment area of upper Karishna river basin and found this distribution best fit for flood prediction at any site of river. Jyothi and Saurabh [22] applied Log Gumbel's Distribution, Log Pearson type 3, Log Normal, Normal and Pearson type 3 distributions and found Log Gumbel's distribution as a best fit for River Kosi. Manas [23] analyzed river Subernarekha using Gumbel Extreme Distribution (GEV) for frequency analysis. Masum [24] compared Powell, Gumbel and Stochastic methods and recommended Gumbel as a best fit for Hydraulic Structures using data of thirteen years.

Luna and Zahid [25] determined the frequency analysis for River Jiya Dhol and resulted that Gumbel Extreme Value and Log Pearson Type 3 are the best fit distributions for this river on the basis of forty years discharge data. Todorovic and Rousselle [26] concluded Gumbel distribution as a best fit approach for the estimation of maximum type events. If data, in form of instant annual flows then Log Normal distribution type 3 was considered as best fit for modeling. Log Pearson 3 and Gumble Extreme Value (GEV) distribution both are considered good for the estimation of high discharge if sufficient data is available.

Ishfaq and Abbas [27] applied generalized logistic, Generalized Extreme value and Generalized Pareto distributions on the data of forty four metrological stations of Pakistan and concluded Generalized Pareto the best one distribution for extreme rainfall assessment. Zakaullah and Mazhar [28] concluded Gumbel distribution best fit approach for the frequency analysis of flood with Chi Square test base after the application of Log Pearson type3, Gumbel, Log Normal and Pearson type 3. Sathe et al., [21] applied Log Pearson Type 3 on the catchment area of upper Karishna river basin and found this distribution the best fit for flood prediction at any site of river. Manas [23] analyzed River Subernarekha using Gumbel Extreme Distribution (GEV) for frequency analysis. Ghorbani et al., [29] utilized various distributions to analyze discharge related flood frequency for Iran. By using the flood index approach, the designed flood quantile gives better estimation for lower return periods i.e 2 and 5 years, while direct interpolation method for higher return periods[30].

Jacob and Osadolor [31] used Extreme value type 1, Log Normal and Log Pearson type distributions on Oshun River, Nigeria and concluded Log Normal distribution as a best fit for return period greater than 50 years and Extreme Value distribution for lesser than 50 years return periods.

The highest flood ever recorded in the Swat River was the mighty flood of July, 2010. Unprecedented monsoon rainfall from 27th July to 30th July 2010 produced flash floods in River Kabul and Swat and caused physical destruction of sixty five communication networks. About 200 millimeter rain fell in 24 hours at many places of Khyber Pakhtunkhwa, 274 millimeters record breaking rain fell within 24 hours in Peshawar. A report published by Federal Flood Commission revealed that 2010 flood bettered the previous records and destroyed Munda and Amandara Headworks, both on Swat River (Annual Flood report 2010", Federal flood commission MOW&P)[32]. During the 2010 flood, the recorded flow of River Swat was 360,000 Cusecs in comparison with normal average flow of 22,557.66 Cusecs in the same month [33].

The flood peak in 2010 at Munda Headworks, River Swat was much higher than the historical peak (with 100-year return period). The flood peak of 8,495 cumec was observed, which was almost 71% higher than its design discharge capacity of 4,955 cumec. Upstream at the Amandara Headworks, in River Swat, a flood peak of 7,646 cumec was observed, which was about 60% higher than the design capacity (4,813 cumec). This flooding at Munda and Amandara headworks was unprecedented, which severely damaged the Amandara Headworks and washed away the Munda Headworks altogether.

Destructive magnitudes of 2010 flood compelled for assessment of flood risk in the Swat valley, so that its vulnerability to flood of such magnitude may be averted. In this study flood frequency analysis against various return periods was carried out using prominent probability distributions and renowned significance tests were applied to suggest the most reliable probability distribution for the study area.

2. Methodology

2.1 Study Area

Swat valley is a central part of significant and strategic regions where three major parts of Asian continent Central Asia, China and south Asia meets. Swat valley is located in North of Pakistan, between 34˚-35˚N latitudes and 72˚-74˚6 E longitude and is spread over an area of about 5737km² with 850 meter average elevation above mean sea level. The Valley physiology changes from South to North with an increase in elevation. River Swat flows along the valley axis. The area is humid with trivial summer season and annually mean rainfall surpasses 1000 millimeters. Annual mean temperature ranges between 18˚C to 30˚C in summer.

Swat River a major tributary of the Indus Basin River system, instigates from Swat-Kohistan region in Kalam and carries the flow of Utror and Ushu rivers and goes downstream in a narrowed ravine up to Baghdheri. The main river (Swat) is 250 Kilometers long and flows are also contributed by number of small and large tributaries namely Adinzai, Swat Ranrizai, Marghazar, Jambil, Arnowai, Barwai, Chail, Daral, Mankial and Gahil. River Swat is Perennial River and is fettered by tall mountains, has diverted its water to thriving monster to release from shackling. The bed slopes of the river vary from 21m/km at Kalam to 4m/km at Munda Headwork.

Hindukush Mountains are the basic source from where glacial water fed the river Swat round the year. It then passes through the Kalam Valley with rushing speed in a constricted gorge (average width of 35 to 40 meters) up-to Madyan. In Upper swat valley, River passes through Narrow valley and Swat-Kohistan is Northern mountainous slice of the valley where forests (alpine) in snow-capped mountains are in abundance. Mankial and Falaksir are the prominent peaks in these mountains. Then the lower plain zones of Swat valley up-to Chakdara about 160 kms. River meets with Panjkora River by entering through narrow gorge in extreme south of the valley and finally drains in to River Kabul near Charsada.

The Swat River basin is predominantly snow-fed and the annual contribution of snowmelt runoff to the total runoff may range from 65–75 %. Snow cover varies throughout the year, which imparts significant impact on snow melt runoff and river discharge. Most of the snow falls usually from November to February since its start from September or October. Snow cover Increases from less than 2 % of the Basin area in August, only at higher altitudes, to about 64 % by the end of January or early February. Snowmelt continues throughout the year but in winter snowmelt runoff remains very low [34]. The contribution of snowmelt, average ground water and average rainfall in the basin is 65%, 19% and 16% respectively [35]. The dominant source of rainfall is westerlies and monsoon which contribute 45% and 55% respectively [36]. The most of the run-off results during spring season due to rapid rise in temperature when the snows on the mountains melt. The rising of river discharge starts from late February and attains highest discharge level during June & July due to supplementation of snow melt water by Monsoon rains. Decline in flow discharge continues till end of October -January, interrupted by occasional floods due to irregular rains [37]. The drainage pattern of the Basin is dendritic due to V shaped valley and the drain area comprises rugged mountains varying in heights from 600m to more than 6,000m with a steady general rise from south to north. The Study location selected herein for purpose of flood frequency analysis is Chakdara Station at River Swat, shown in Figure 1.

2.2 Data Set

Collection of needed data is one of the most important processes for analysis. The data is essential in the application of different empirical approaches. The Stream Flow/ Discharge daily data of 58 years i.e from year 1961 to 2018 of River Swat recorded at Chakdara station was obtained from WAPDA. The stream gauging station i.e Chakdara Station installed in 1960, having catchment area of 5776 Km2 and elevation of 676 meter, managed by Surface Water and Hydrology (SWHP), WAPDA. The latitude and longitude of the Chakdara guage station are 34o 38/ 40// and 72o 01/ 30// respectively. The Bar graph representing peak annual discharge of 58 years from 1960 to 2018 recorded at River Swat (Chakdara Station) is shown in Fig. 2. The peak annual flow data recorded at Chakdara Station at river Swat was used for flood frequency analysis.

2.3 Flood Frequency Analysis

Flood Frequency Analysis (FFA) is the computation of interrelation between Relative Frequencies (RF), Probability of Exceedance (POE) and Average Reoccurrence Interval (ARI). Where POE is probability of flow equal or greater than a specific value and ARI is the return period, which can be described as:

$$\text{POE} = \text{P(Z)} \tag{1}$$

$$\text{RF} = \text{F(Z)} \tag{2}$$

$$\text{ARI} = \text{T(Z)} \tag{3}$$

Where:

$$\text{P(Z)} = 1 - \text{F(Z)} \tag{4}$$

$$\text{T(Z)} = \frac{1}{\text{P(Z)}} = \frac{1}{1-\text{F(Z)}} \tag{5}$$

Peak Annual discharge data is ranked either from lowest to highest peak value or vice versa and calculations are made for F (Z) as under:

$$\mathbf{F(Z)} = \frac{\mathbf{n+1}}{\mathbf{R}} \tag{6}$$

R is rank of a single flood event, Z from the selected data series and n is size of series. In flood frequency analysis the annual peak discharge in a river is fitted in different probability distributions. Five commonly used distributions in flood frequency analysis, Normal Distribution, Log-Normal Distribution, Gumbel (Extreme Value Type 1) Distribution and Log-Pearson Distribution are used in this study against 2, 5, 10, 20, 25, 50, 100, 500, 1000 and 10000 year return period.

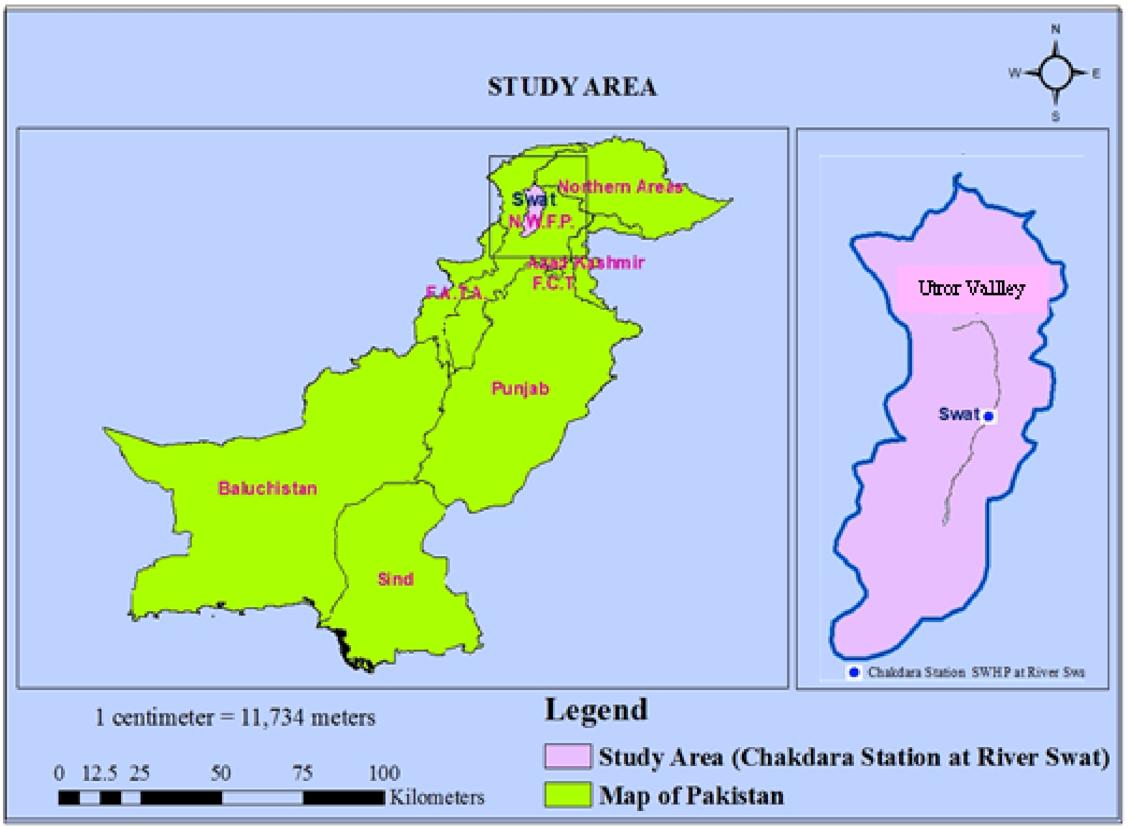

Fig. 1: Study area and location of Chakdra gauging station.

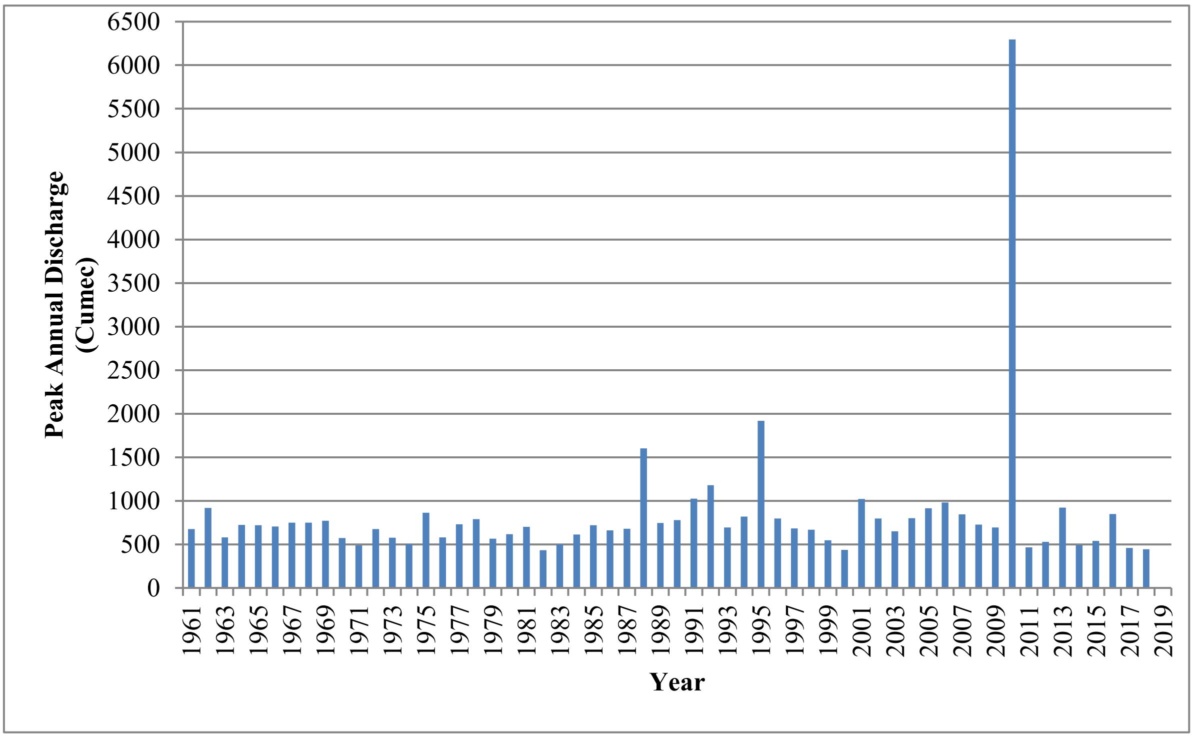

Fig. 2: Peak Annual Discharge Recorded at River Swat (Chakdara Station)

2.4 Probability Density Function (PDF)

Probability-density function is evaluated to determine probability that a variable has value of x. It is a relationship between a variable taken on X-axis and the respective probabilities on Y-axis. It is impossible to determine the exact probability against a value, because it may be zero. Therefore probability is determined with some tolerance if c and d are the maximum and minimum tolerance then PDF is evaluated by using Equation 7. The area under the curve is determined from c to d and the whole area under the curve should be equal to 1.

$$\int_c^d f(x)dx = P(c \leq x \leq d) \quad (7)$$

2.5 Cumulative-Distribution Function (CDF)

Cumulative-distribution function (CDF) is plotted to determine probability of a variable equal or less than x. It is a relationship between a variable taken on X-axis and the respective probabilities on Y-axis. The empirical (observed) CDF is plotted in the form of stepped dis-continuous line, depending upon numbers of bin and theoretical CDF for various distributions as continuous curve for this research case CDF is determined by using equation given below:

$$F(x) = \int_{-\infty}^{x} f(u)du \quad (8)$$

2.6 Probability-Probability Plot (P-P Plot)

The P-P plot is graph between probabilities of theoretical (distributional) and observed (input) data values. Such graphs define points at which data is following and the points at which data is not following a theoretical distribution. The Probabilities of observed (empirical) data set is taken on horizontal axis while the theoretical (distributional) on vertical axis. A reference line at 45^o is taken to check the departure of data sets from each other, departure from reference line indicates failure of a distribution for the data set. For correct modelling the plot of specified distribution should be approximately linear.

2.7 Quantile-Quantile Plot (Q-Q Plot)

The Q-Q plot is a graph between quantiles of distributional (Theoretical) data set values and observed (input) data set values. The empirical (observed) data set is taken on Horizontal axis while the distributional (Theoretical) one on Vertical axis. A reference line at 45^o is taken to check the departure of data sets from each other, greater the departure from reference line indicates failure of a distribution for the data set. For correct modelling the plot of specified distribution should be approximately linear, given by Equation 9.

$$F^{-1}\left[F_n(x_i) - \frac{0.5}{n}\right] \quad (9)$$

2.8 Probability Difference Graph (PDG)

It is a plot between empirical and theoretical CDF. The plot determines well fit theoretical distribution against observed data. It also compares Goodness of Fit (GOF) of numerous fitted distributions and is displayed by scatterplot. The observed data is taken on X-axis while the difference in between Empirical Cumulative Distribution Function (ECDF) and Theoretical Cumulative Distribution Function (TCDF) on Y-axis, given by Equation 10.

$$Diff(x) = F_n(x) - F(x) \quad (10)$$

2.9 Significance Test

To measure compatibility of a random sample from observed data with theoretical Probability Distribution Function (PDF) tests are performed called Goodness of Fit (GOF). Defining test statistic, (function of data) by measuring distance between data and hypothesis and calculation of probability of the obtained data has larger value than observed value is general procedure behind these tests. Small probabilities (<1%) indicate poor fit, while high probabilities (≈1%) corresponds to best fit. Normally Kolmogrov Smirnov, Anderson darling and Chi-squared tests are used for the purpose of GOF. Kolmogrov Smirnov test rely on function of cumulative distribution (ECDF) and decides whether a sample is from hypothesized continuous distribution (HCD) or not. The Anderson darling test makes comparison between Observed Cumulative Distribution Function (OCDF) and Expected Cumulative Distribution Function (ECDF), but in comparison to Kolmogorov Smirnov, it gives larger weight to tails. The Chi-squared test is for scattered data and test statistic value depends on scatterings of data and is utilized to know if sample from population comes with certain distribution.

In order to determine the most applicable Probability Distribution, *Easy Fit Model was used*. *Easy Fit* is an application for data analysis and simulation, allowing fitting the probability distributions into sample data. It is the combination of classical statistical analysis methods & advanced data analysis techniques which helps in taking decision regarding probability.

3. Results and Discussion

The results are based on Peak annual stream flow data recorded at Wapda stream gauging Station i.e. Chakdara Station, River Swat. The obtained data of 58 years i.e. from 1961 to 2018 taken from SWHP WAPDA was evaluated for the purpose of flood frequency analysis using Normal, Log-Normal, Extreme Value (Type 1) Gumbel, and Log-Pearson-3 distributions. The discharges were estimated for 2, 5, 10, 20, 25, 50, 100, 500, 1000, and 10000 year return period (Table 1)

3.1 Probability Density Function (PDF)

Discharge values (Cumec) of Chakdara gauge Station at River Swat recorded by SWHP WAPDA and distributional data are taken on X-axis while the respective probabilities on Y-axis. The empirical (observed) PDF was plotted as histogram, which consisted of vertical bars [bin].

Each Vertical bar defines sample data falling within the respective interval. While the theoretical PDF are displayed as continuous scaled curve for various distributions on the basis of numbers of interval. Scale means multiplication of PDF values with interval width. The results of PDF Plots of specified distributions obtained from simulation of Easy-fit Model are shown in Fig. 3.

The results obtained from Easyfit simulation model depicts that Extreme Value (Type-1) distribution follows the observed data record, initiates from Zero probability for nil discharge and limb fluctuates with increase and decrease of observed data set. The distribution follows the observed data record for high discharges but not for very low discharges. While the log normal distribution computes 0.98 probability for nil & very low discharge record but approximately follows the observed data record for very high discharges. Discussing about log Pearson 3, the probability is more than one for nil discharge. So the estimation through log Pearson 3 is an over estimate. The probability of both Gumbel and Normal distribution is less than the observational probability. The distributional data neither achieves peak with respect to observed data set for low discharges nor follows the observed data set for very high discharges. So the estimation using both Gumbel and Normal distribution is an under estimate.

Therefore, it is concluded that as far as, the Probability-Density Function is concerned, the Gen Extreme Value [type I] distribution is comparatively better than other applied distribution in the study.

Table 1: Estimated Discharges Using Different Statistical Distributions

Return Period (Year)	Normal Distribution (m^3/sec)	Log Normal Distribution (m^3/sec)	Extreme Value Type I (m^3/sec)	Log Pearson III (m^3/sec)	Gumbel (m^3/sec)
2	830.92	729.05	703.87	600.14	703.87
5	1481.72	1027.87	1387.36	846.19	1387.36
10	1822.23	1230.24	1839.89	1173.06	1839.89
20	2103.35	1427.01	2273.97	1682.31	2273.97
25	2185.23	1490.04	2411.66	1900.34	2411.66
50	2419.66	1686.29	2835.84	2817.87	2835.84
100	2630.49	1884.78	3256.88	4265.50	3256.88
500	3057.20	2360.86	4229.84	11944.00	4229.84
1000	3221.18	2574.29	4648.13	19086.02	4648.13
10000	3707.35	3327.35	6036.92	99463.80	6036.92

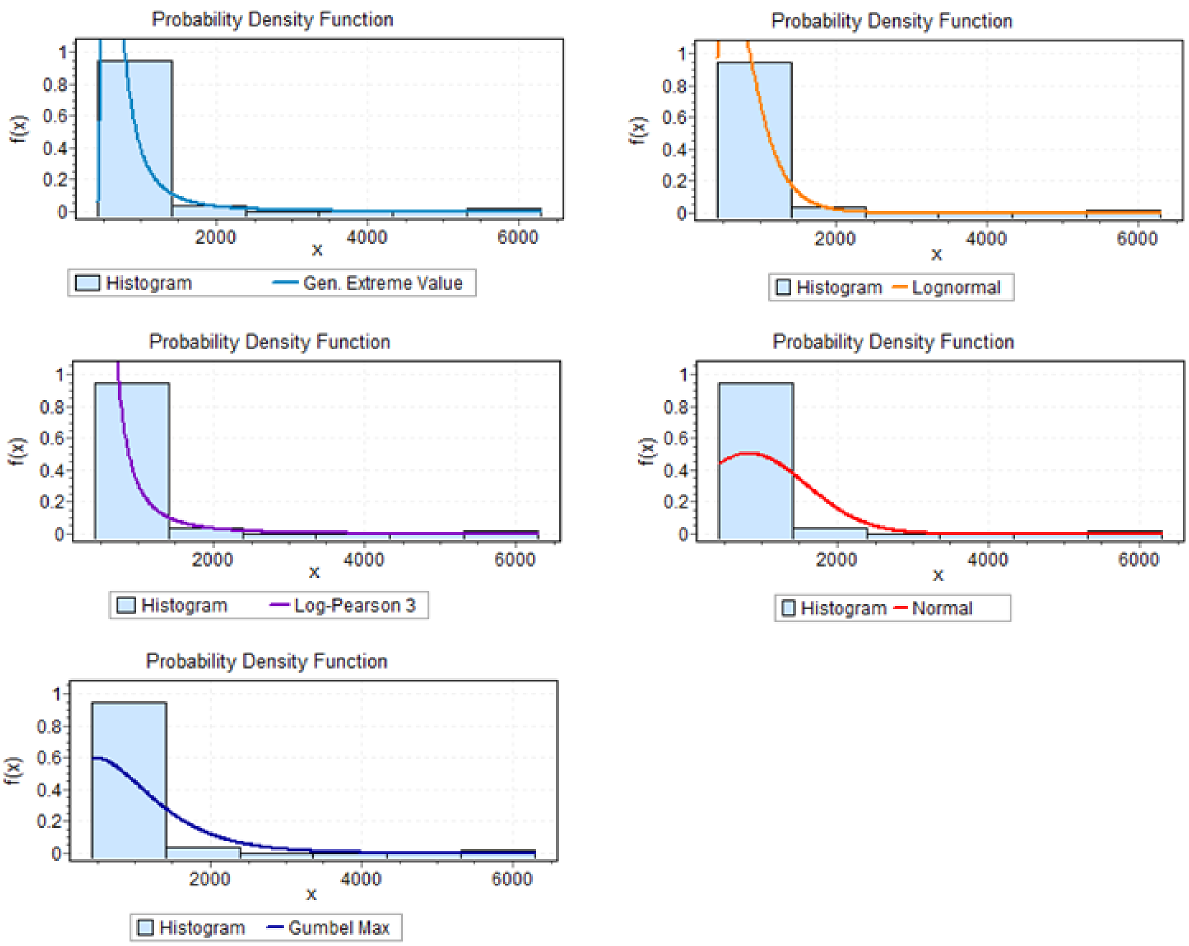

Fig. 3: PDF Plots of Different Statistical Distributions (x = discharge in m^3/sec, f(x) respective probability)

3.2 Cumulative Distribution Function (CDF)

The results obtained from Easyfit simulation model shown in Fig. 4 depicts that Gen Extreme Value [Type I] distribution obeys probability of observed data record of discharges. For very high and medium discharges using Gen Extreme Value [Type I] distribution, the probability of measuring CDF approximately follows empirical CDF and fits best for high and medium discharge values. But for low discharges, the probability is less than the empirical one ranging between Zero and 1. The log Normal distribution did not follow the probability of observed data record of discharges. The probability is less than the empirical one ranging between Zero and 1 for low discharges, while for medium and high discharges the probability is more than 1, which means probability of measuring CDF is more than the empirical CDF. So the estimation using log normal distribution is an over estimate. The log Pearson 3 distribution did not estimate discharge rate less than 550 m^3/s and the probability is less than the empirical one ranging between Zero and 1 for low discharges. However, follows observed discharge data record for medium and high discharges i.e. the probability of measuring CDF approximately follows empirical CDF. Gumbel and Normal distribution estimates less than the empirical one ranging between Zero and 1 for low and medium discharges, while for high discharges the probability is more than 1, which means probability of measuring CDF is more than the empirical CDF.

Therefore, it was concluded that as far as, the Cumulative distribution function is concerned, as comparatively probability of measuring CDF using Gen Extreme Value [type I] distribution follows empirical CDF better than others. Therefore, is comparatively better than other applied distribution in the study

3.3 Probability-Probability Plot (P-P Plot)

The results of P-P plot obtained from Easyfit simulation model are shown in Fig. 5.The result of Log Normal distribution depicts that reference line is very far from probability of distribution. Moreover, it revealed higher probability against lower discharge values and very low probability against higher discharges as compared to the observed probability. Log Pearson 3 distribution doesn't estimate the probability of low discharges, estimated higher probability for medium discharges and lower probability against higher discharges as compared to observed probability. While both Gumbel and Normal distributions estimated higher probability for lower discharges and low probability for high discharge values. Gen Extreme Value [Type I] distribution also estimated high probability for medium discharge values but approximately showed linear plot for low and high discharge values, but comparatively better than the others.

Thus from the results obtained from the simulation of Easyfit model, it is concluded that Extreme Value [Type I] distribution is comparatively better than the other applied distributions.

3.4 Quantile-Quantile Plot (Q-Q Plot)

The results obtained from Easyfit simulation model shown in Fig. 6, depicts departure between reference line and quantiles of distribution. The discharge data set didn't follows the reference line in case of Normal, Gumbel and log pearson 3 distributions and didn't consider discharge less than 550 Cumec. In case of Normal and Gumbel distribution, measured discharge values were very far from reference line for low and very high discharge values. The variations in estimated discharge values were observed in form of overestimation at some places while underestimation at the others. Log Pearson 3 didn't consider very low discharge values though the reference line was followed at low discharges. Furthermore, measured discharge values were very far from reference line for high discharge values. In case of Gen Extreme Value [Type I] distribution, departure from reference line at some points as positive and at some points as negative was observed. The distribution follows linear relationship for low to medium discharge values except for high discharges.

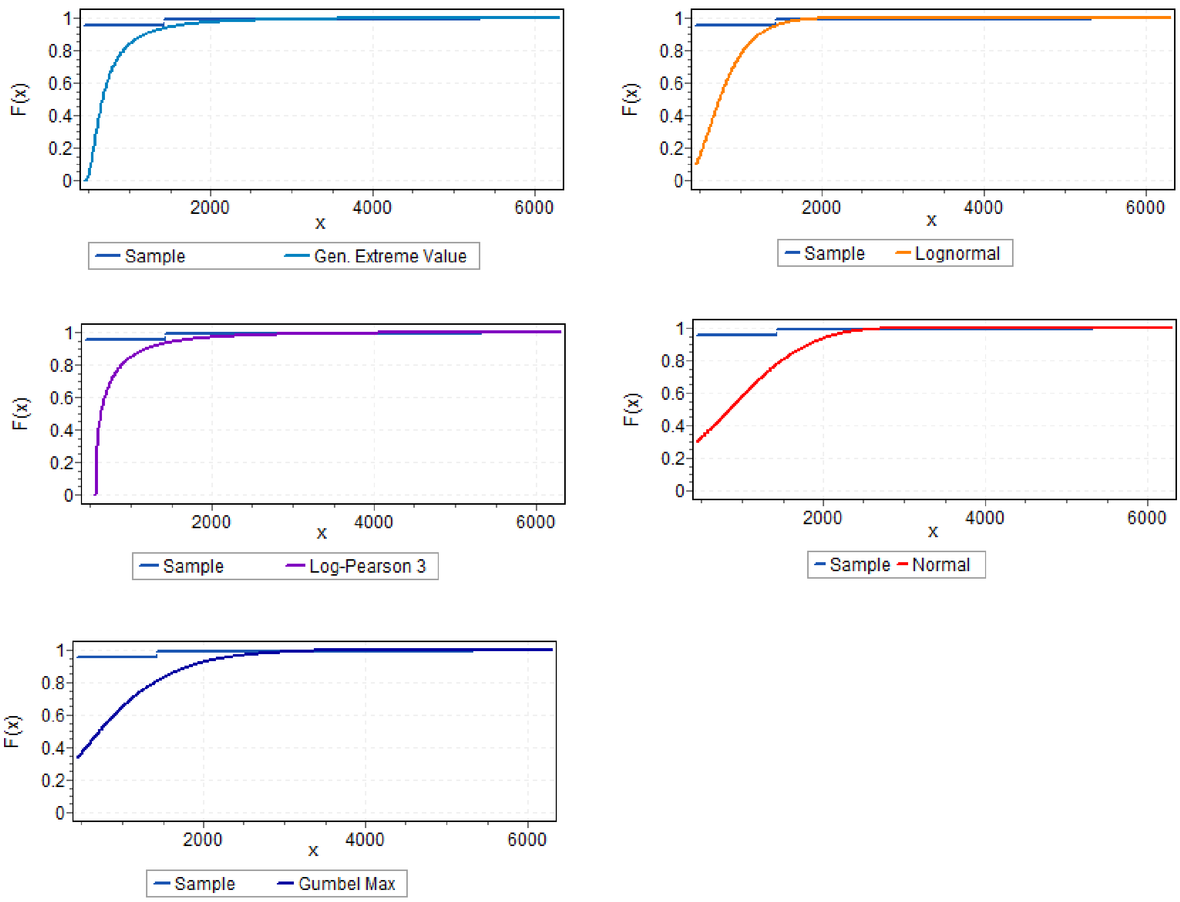

Fig. 4: Cumulative Distribution Function Plot of Different Statistical Distributions

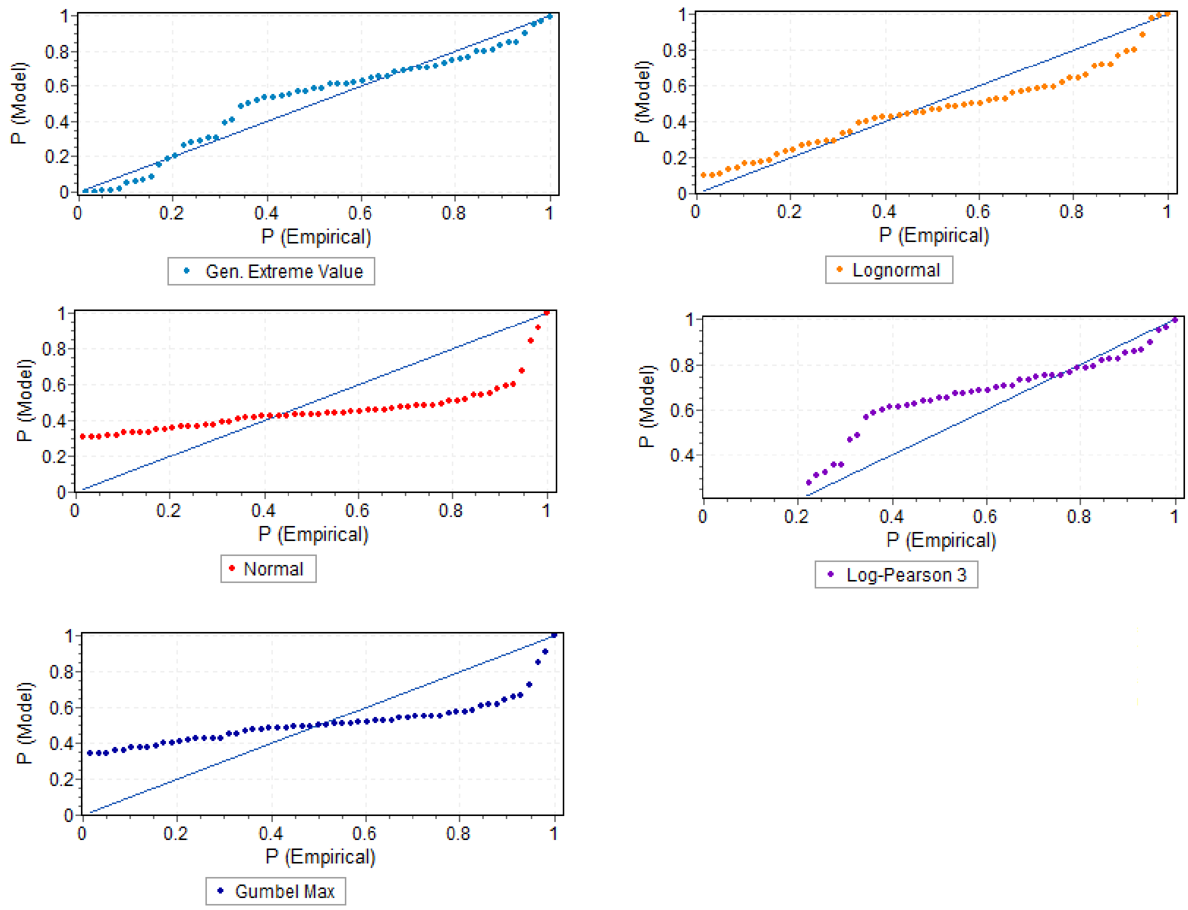

Fig. 5: P-P Plots of Different Statistical Distributions

The Q-Q plots of Normal, Log Normal, Gumbel and Log Pearson 3 distributions indicated more departure from reference line as compared to Extreme Value [Type I] distribution, So it was concluded that the plot of Gen Extreme Value [Type I] fits best.

3.5 Probability Difference Graph (PDG)

The results obtained from Easyfit simulation model are shown in Fig. 7, which depicts that in case of Normal, Log Normal, Gumbel and Log Pearson 3 distributions, for very low discharges negative difference (Under-estimates) by comparing with referred line while positive difference for medium discharges (Over-Estimates) and an accurate measure for high discharges were observed. Extreme Value [Type I] distribution showed departure from reference line at some points as positive and at some points as negative, positive difference for very low and medium discharge up to 0.075 and negative difference up to 0.105 and followed the trend for very higher discharges.

Thus from the results of Easyfit model, it was concluded that Gen Extreme Value [Type I] distribution is comparatively better than the other selected distributions.

3.6 Significance Test

In order to determine the best fit distribution among Lognormal, Normal, Gumbel, Log-Pearson 3 and Extreme Value (Type-1), Peak Annual flow data of 58 years i.e from year 1961 to 2018 of River Swat recorded at Chakdara station and three significance tests namely Anderson Darling, Kolmogrov Smirnov and Chi-Squared were applied for the selection of a Probability Distribution using Easyfit Model. The Results are shown in Table 2.

It was concluded that Gen Extreme Value [Type I] was the most applicable Probability Distribution among the other specified distribution for the Study area. The extreme value (Type-1) ranked first by applied significance tests i.e. Kolmogrov Smirnov, Anderson Darling and Chi-Squared.

Table 2: Summary of Tests of Goodness fit

Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
	Statistic	**Rank**	**Statistic**	**Rank**	**Statistic**	**Rank**
Extreme Value Type I	0.1575	1	1.9503	1	2.8576	1
Log normal	0.16775	2	2.5975	2	13.742	2
Log-Pearson 3	0.23839	3	43.887	5	-	-
Normal	0.33295	4	13.12	4	66.012	3
Gumbel	0.33766	5	10.02	3	67.712	4

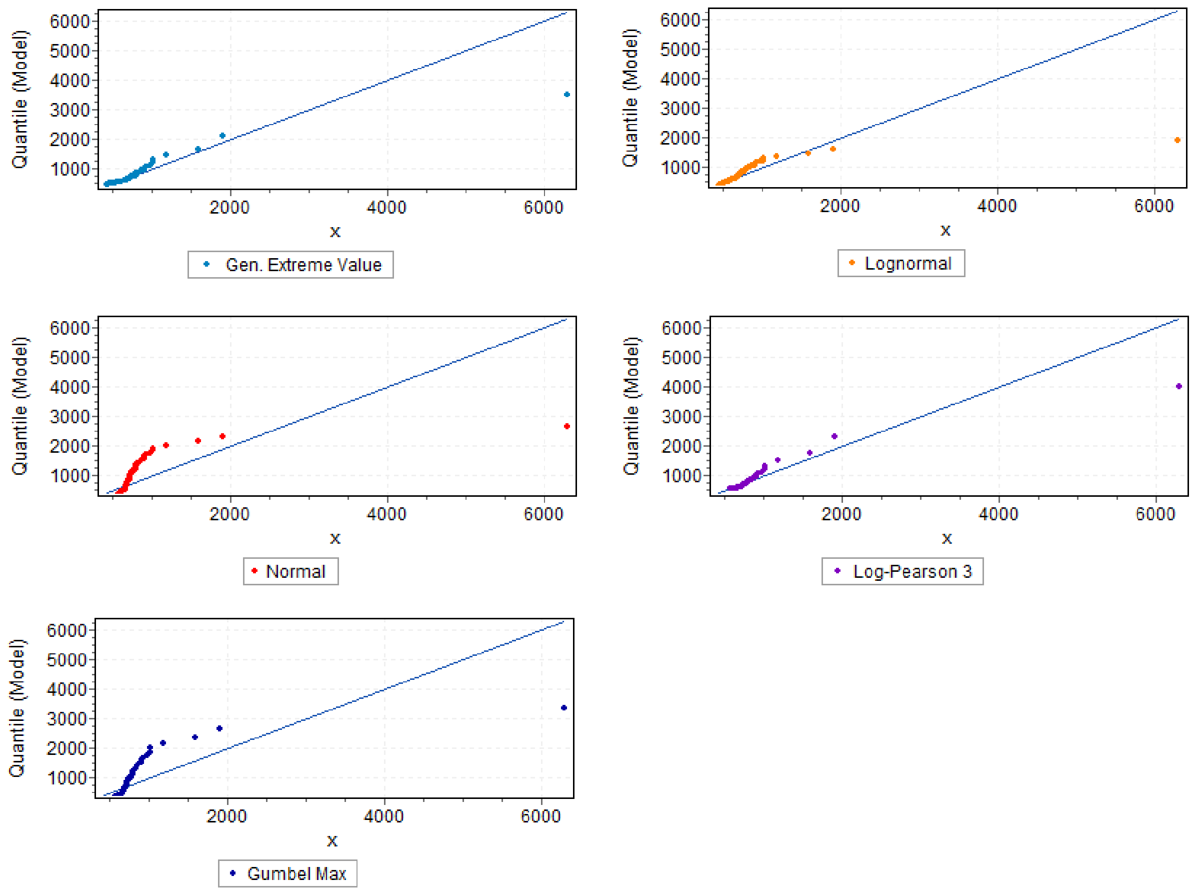

Fig. 6: Q-Q Plots of Different Statistical Distributions

4. CONCLUSIONS

In this study flood frequency analysis of river Swat was carried out at Chakdara Station using five probability distributions i.e. Gumbel, Extreme Value (Type-1), Normal, Log Normal and Log Pearson Type III distributions against 2, 5, 10, 20, 25, 50, 100, 500, 1000, and 10000 year return periods to assess the occurrence of flood events like 2010. Since the catchment area of river swat is hilly undulated terrain having poor communication network and difficult accessibility, where precipitation and flow gauging stations are sparse.

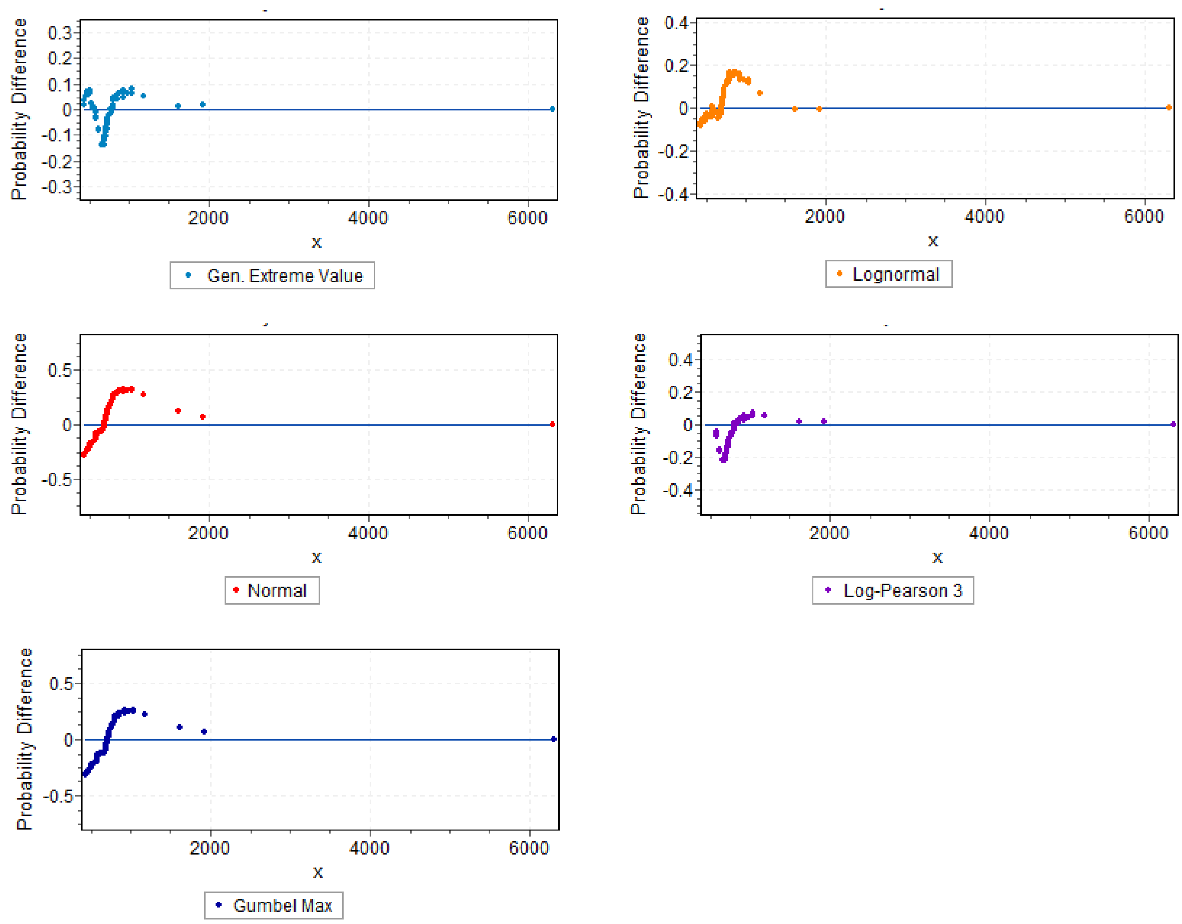

Fig. 7: PDG Plots of Different Statistical Distributions

In this study flood frequency analysis of river Swat was carried out at Chakdara Station using five probability distributions i.e. Gumbel, Extreme Value (Type-1), Normal, Log Normal and Log Pearson Type III distributions against 2, 5, 10, 20, 25, 50, 100, 500, 1000, and 10000 year return periods to assess the occurrence of flood events like 2010. Since the catchment area of river swat is hilly undulated terrain having poor communication network and difficult accessibility, where precipitation and flow gauging stations are sparse.

The PDF results revealed that Extreme value (Type-1) distribution followed the observed data record for very high discharges except for very low discharges as compared to other four applied distributions. Similar was the case for Cumulative distribution function. Whereas, comparatively probability of measuring CDF using Gen Extreme Value [type I] distribution followed empirical CDF better than others. Furthermore, from the P-P Plot, it was concluded that though Extreme Value [Type I] distribution estimated high probability for medium discharge values but approximately showed linear plot for low and high discharge values, which was comparatively better than other applied distributions. The Q-Q plots of Normal, Log Normal, Gumbel and Log Pearson 3 distributions indicated more departure from reference line as compared to Extreme Value [Type I] distribution. So the plot of Gen Extreme Value [Type I] fitted best. Extreme Value [Type I] distribution showed departure from reference line in probability density graph i.e. at some points as positive and at some points as negative but followed the trend better as compared to the others.

Significance tests namely Anderson Darling, Kolmogrov smirnov and Chi-Squared were applied for selection of a Probability Distribution using Easyfit Model. It was concluded that Gen Extreme Value [Type I] was the most applicable Probability Distribution among the other specified distribution for the Study area. The extreme value (Type-1) was ranked first by applied significance tests i.e. Kolmogrov smirnov, Anderson Darling and Chi-Squared.

The flood management approach currently in use has no provisions for floods exceeding design limits. Due to changes in the patterns of flooding and in the behavior of streams, the design limits and criteria for Major River structures, as well as structures in rural and urban areas, should be reviewed. For the purpose of Culverts and permanent bridges, 100 year return period flood is preferred to be used as design flow. Out of various probability distributions Extreme Value (Type-1) Distribution was found to be the best fit against reference data. Therefore, the estimated flood at 100 year return period may be used. As The flood peak in 2010 at Munda Headworks, River Swat was much higher than the historical peak (with 100-year return period), therefore for headworks the estimated flood of the best fit distribution i.e. Extreme Value (Type-1) at 500 year return period may be used. For large dams the estimated flood, using the best fit distribution i.e. Extreme Value (Type-1) with 5 year return period for diversion scheme and 10,000 year return period for spillway design may be used.

5. Acknowledgement

The authors are grateful to Surface Water and Hydrology Project (SWHP), WAPDA for providing the flows data used in the study. The authors are grateful to the staff of Centre of Excellence in Water Resources Engineering (CEWRE) for providing technical support on various occasions provided during the completion of the study.

6. References

[1] ADB .2010. Pakistan Floods 2010: Preliminary Damage and Needs Assessment. *Asian Development Bank.*

[2] Cherng, H., et al., 2011. Roles of European blocking and tropical extratropical interaction in the 2010 Pakistan flooding. *Geophysics Research Letters*, 38, 314-329.

[3] Abdurahim, H., et al., 2020. Flood vulnerability assessment using MOVE framework: a case study of the northern part of district Peshawar, Pakistan, Natural Hazards: *Journal of the International Society for the Prevention and Mitigation of Natural Hazards*, vol. 101(2), pages 385-408.

[4] Akhter, A., Dil-Bahadur, R., 2019, Localized Floods, Poverty and Food Security: Empirical Evidence from Rural Pakistan, *Journal of Hydrology*, 7, 2; doi:10.3390/hydrology7010002.

[5] Akiyuki, K., et al., 2020. A local level relationship between floods and poverty: A case in Myanmar, International *Journal of Disaster Risk Reduction*, Vol 42, 101348.

[6] Opere, A., et al., 2006. At site flood frequency analysis for the nile equatorial basins. *Physics and Chemistry of Earth*, 31, 919-927.

[7] Ahmad, B., et al., 2010. Hydrological modelling and flood hazard mapping of Nullah Lai. *Pakistan Academy of Sciences, Proceedings,* 47 (4), 215- 226.

[8] Khaliq, M., et al., 2006. Frequency analysis of a sequence of dependent and /or non-stationary hydro-meteorological observations: A review. *Journal of Hydrology*, 329(3), 534–552.

[9] Law, G.S., and Tasker, G.D., 2003. Flood-frequency prediction methods for unregulated streams of Tennessee. *Water Resources Investigations* Report 03-4176, Nashville, Tennessee.

[10] Robert, M.H., 1987. Plotting positions for historic floods and their precision. Eos, AGU Trans, 23, 715-719.

[11] Moughamian, M.S., et al., 1987. Estimation of flood frequency, an evaluation of two derived distribution procedures. *Water Resources Research*, 23 (7), 1309-1319.

[12] Hromadka, T.V. and Whitley, R.J., 1989. Checking flood frequency curves using rainfall data. *Journal of Hydraulic Engineering,* 115, 544-548.

[13] Burn, D.H. and Goel, N.K., 2001. Flood frequency analysis for the red river at Winnipeg. *Canadian Journal of Engineering*, 28, 355-362.

[14] Aksoy, H., 2000. Use of gamma distribution in hydrological analysis. Turkish *Journal of Engineering and Environmental Sciences,* 24, 419-428.

[15] Lanxin Hu., et al., 2019. Sensitivity of flood frequency analysis to data record, statistical model, and parameter estimation methods: An evaluation over the contiguous United States, *Flood Risk Management,* Vol. 13, Issue-1.

[16] McMahon, T.A. and Srikanthan, R., 1981. Log-Pearson type 3 distribution: is it applicable to flood frequency analysis of Australian streams. *Journal of Hydrology,* 52, 139-147.

[17] Vogel, R.M. and Wilson, I., 1996. Probability distribution of annual maximum, mean, and minimum streamflows in the United States. *Journal of Hydrologic Engineering,* 1(2), 6976.

[18] Sasan, K., et al. 2020, Regional Flood Frequency Analysis Using an Artificial Neural Network Model, *Geosciences* 2020, 10(4), 127; https://doi.org/10.3390/geosciences10040127.

[19] Khan, M.R., 2013. Frequency analysis of flood flow at garudeshwar station in narmada river, Gujarat, India. Universal *Journal of Environmental Research and Technology,* 3, 39-54.

[20] Wisam, A., et al., 2013. Flood Frequency Analysis for Greater – ZAB River. *Journal of Babylon*, 21.

[21] Sathe, B.K., et al., 2012. Rainfall analysis and design flood estimation for upper Krishna River basin catchment in India. *International Journal of Scientific & Engineering Research*, 3 (8), 2229-5518.

[22] Jyothi, P. and Saurabh, S., 2015. Flood frequency analysis of river kosi, Uttarakhand, India using statistical approach. *International Journal of Research in Engineering and Technology*, 4 (8), 25-39.

[23] Manas, K.M., 2013. Flood frequency analysis of river Subernarekha, India, using gumbel's extreme value distribution. *International Journal of Computational Engineering Research*, 3 (7), 126-138.

[24] Masum, A.A., et al., 2014. Flood frequency analysis by probability and stochastic method for Padma river, Bangladesh. *American Journal of Civil Engineering.* 2 (1), 8-11.

[25] Luna, M.D. and Zahid, H.Q., 2014. Flood frequency analysis for Jiya Dhol river of Brahmaputra Valley. *International Journal of Sciences Basic and Applied Research*, 14 (2), 14-24.

[26] Todorovic, P. and Rousselle, J., 1971. Some Problems of Flood Analysis. *Water Resources Research*, 7(5), 1144-1150.

[27] Ishfaq, A. and Abbas, A., 2015. Total Annual Rainfall Frequency Analysis in Pakistan Using Methods of L-Moments and Tl-Moment *Science Int*.,27(3),2331-2337.

[28] Zakaullah, and Mazhar, M.S., 2012. Flood frequency analysis of homogeneous regions of Jhelum River Basin *International Journal of Water Resources and Environmental Engineering*. Vol. 4(5), pp. 144-149.

[29] Ghorbani, A., et al., 2010. Flood frequency analysis using Mathematica *Turkish J. Eng. Env. Sci*.34, 171–188.

[30] S. Baidya., et al., 2020. "Flood frequency analysis," Natural Hazards: Journal of the International Society for the Prevention and Mitigation of Natural Hazards, *Journal of the International Society for the Prevention and Mitigation of Natural Hazards*, vol. 100(3), pages 1137-1158, February.

[31] Jacob, O.E. and Osadolor, C.I., 2013. Flood Frequency Analysis at Oshun River in Asejire Dam Site, Nigeria. *Journal of Earth Science and Engineering*, 3 (13), 292-300.

[32] FFC. 2010. "Annual Flood report 2010" *Federal Flood Commission Ministry of Water & Power Pakistan.*

[33] A.M. Yousafzai. et.al., 2013. Fresh Records on Water Quality and Ichthyodiversity of River Swat at Charsadda, Khyber Pakhtunkhwa. *Pakistan J. Zool*., vol. 45(6), pp. 1727-1734, 2013.

[34] Zakir.H.D., et.al. 2011. Satellite-Based Snowcover Distribution and Associated Snowmelt Runoff Modeling in Swat River Basin of Pakistan. *Proceedings of the Pakistan Academy of Sciences* 48 (1): 19–32, 2011.

[35] Muhammad, U., et.al, 2015. Estimation of Snowmelt Contribution for Kalam Catchment. *Pak. J. Engg. & Appl. Sci.* Vol. 17, Jul., 2015 (p. 64–79).

[36] Malik, M.A., and Bhatti, A.Z. 2015., Characterizing Snowmelt Regime of the River Swat - A Case Study. *Technical Journal, University of Engineering and Technology (UET) Taxila*, Pakistan Vol. 20 No. II-2015.

[37] Bibi, T., et.al., 2018. Flood risk assessment of river Kabul and swat catchment area: district Charsadda, Pakistan. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* Volume XLII-4/W9. Kuala Lumpur, Malaysia.