Pol. J. Environ. Stud. Vol. XX, No. X (XXXX), 1-12
DOI: *10.15244/pjoes/205074* ONLINE PUBLICATION DATE: 2025-07-14

Original Research

At-Site Flood Frequency Analysis: Evaluating Some Candidate Probability Models

Sajjad Haider Bhatti[1]*, Muhammad Irfan[2], Sandile Charles Shongwe[3], Muhammad Ali Raza[2]

[1]College of Statistical Sciences, University of the Punjab, Lahore, Pakistan
[2]Department of Statistics, Government College University, Faisalabad, Pakistan
[3]Department of Mathematical Statistics and Actuarial Science, University of the Free State, Bloemfontein, South Africa

Received: 21 January 2025
Accepted: 13 May 2025

Abstract

The article presents flood frequency modeling of the Sutlej River's gauging sites located in Pakistan. The objective is to search for the most cogent mix of probability distribution and parameter estimation approach for each site separately. Different probability distributions that have been recommended in the literature are fitted for modeling maximum annual water discharge. Separately for each site, the parameters of the competing probability distributions are estimated through different estimation strategies. Different compounds of distributions and estimation methods are compared for each site through seven different performance metrics, and the most plausible compound is recommended for that particular gauging station. Results from ranking different models revealed that the generalized Pareto distribution estimated through the weighted least squares method is most appropriate for both sites considered for analysis. Finally, based on the selected blend of distribution and estimation method, an estimate of annual maximum flow, along with 95% parametric bootstrap confidence intervals, is provided for different return years.

Keywords: flood frequency analysis, extreme value models, least squares, L-moments, maximum likelihood, Sutlej River

Introduction

Floods are the most destructive natural disasters that cause losses of lives, land, infrastructure, and agricultural and industrial production, and hence are considered to be one of the biggest threats to the socio-economic fabric in affected areas. It can also create adverse social and environmental problems like disturbance in electricity supply and associated risks, contamination in drinking water, sanitary and drainage hazards, and destruction in communication networks [1-4]. A flood is actually a phenomenon representing the overflow of surplus water from a river, lake, or any other water channel to land outside the usual path of that river or lake. Floods are primarily caused by melting glaciers, extreme rains, leakage or breakage of dams or other water storage infrastructures, or the incapacity of rivers or water streams to pass the sudden additional amount of water [5-10]. Floods are natural disasters that recur over time with irregular

*e-mail: sajjad.stat@pu.edu.pk

intervals. It is a continuous hazardous phenomenon that has affected humans across different time periods. Flood frequency analysis (FFA) is an important way of predicting the time of occurrence of certain levels of floods in the future, which is necessary for planning and designing water reservoirs, disaster management, and other related activities [3, 11-14]. In the literature, different approaches (like California, Gumbel, and Hazen) are recommended for analyzing the frequency of floods [5, 15]. For modeling the stream flow or flood frequency, there are different approaches, like at-site, regional (considering different gauging stations as a homogeneous region), or country-level (one model for all sites in the country), that can be applied. The at-site FFA is the simplest and most direct method for predicting flood magnitude and its recurrence time period with a certain probability for any particular station of any river [2, 9, 10, 16, 17]. A regional approach for FFA is a useful option if data from individual gauging stations is not available for long time periods [18, 19].

For at-site FFA, the data series is assumed to follow a parent probability distribution, and then the parameter estimates of that distribution are used to predict certain flood magnitudes in the future. The optimal choice of distribution and estimation strategy is pivotal in achieving more accurate flood predictions, but there is no general agreement in the literature about any single probability distribution or any particular estimation method [3, 11, 20]. Each probability model has its own advantages and disadvantages in modeling any particular station's flood data, and different models can be suitable for different stations. For FFA of many rivers in different countries, varying probability distributions have been recommended, including (but not limited to) normal, exponential, generalized normal, two-parameter lognormal, three-parameter lognormal, four-parameter Wakeby, five-parameter Wakeby, two-parameter gamma, three-parameter gamma, generalized logistic, Pearson type-III, generalized extreme value, Gumbel, and generalized Pareto [5, 9, 10, 21-23]. Some very informative and comprehensive reviews on the collection of probability distributions used for FFA in different countries are available in the literature [2, 24-30].

For at-site modeling of annual peak discharge at two gauging stations of the Sutlej River, we have applied Generalized Logistic (GLO), Gumbel (GUM), Reverse Gumbel (REV-GUM), Generalized Pareto (GP), and Generalized Extreme Value (GEV) as candidate probability distributions. These distributions have been applied and recommended for FFA in different studies focusing on different rivers in different countries around the globe [5, 10, 21, 31-35].

Different approaches for the parameter estimation of probability distributions are available in the literature. However, as far as the distributions used in FFA are concerned, methods of L-moments and maximum likelihood are applied in most cases. The maximum likelihood method is one of the widely applied estimation methods, while the L-moments method recently attained widespread attention, particularly in hydrological studies, due to its computational simplicity [3, 9, 10, 21, 36]. However, the least squares method and its variants also provided efficient estimation and were recommended in different cases [37]. Therefore, in the current work, we employed various estimation methods for estimating the parameters of candidate probability models. These methods are L-Moments (LM), Maximum Likelihood (ML), Least Squares (LS), Weighted Least Squares (WLS), and Relative Least Squares (RLS).

The article aims to find the optimal mixture of distribution and estimation method for modeling yearly extreme flow data taken from different gauging sites located along the Sutlej River separately. The performance of each amalgamation (mixture of distribution and estimation method) is assessed based on different performance metrics. For each gauging site, we have evaluated 25 models (5 probability distributions, each estimated with 5 different estimation methods).

Once the best-fit model (blend of distribution and estimation method) for a particular site is identified, it is used to estimate annual extreme discharges for many return years (like 5, 10, 25, 50, 100, 200, 500, and 1000) with a certain non-exceedance probability.

The remaining part of the article is organized as follows: Section 2 gives an overview of the locations of the Sutlej River and data utilized in this study; Section 3 outlines the candidate probability distributions and methods for the parameter estimation applied in this work; Section 4 provides the results and comparative analysis of applying various mixtures of probability distributions and estimation techniques; and Section 5 summarizes the significant findings and conclusions that can be inferred from the current analysis.

Study Locations and Data

The Sutlej River is a key waterway that plays a vital role in agriculture and other related activities in the southern part of Pakistan's Punjab province. The river flows northeast to southeast, entering Pakistan from the Indian Punjab. In Pakistan, it covers a distance of around 350 km before merging into the Chenab River at Panjnad. Fig. 1 represents a map of the Sutlej River with gauging sites in Pakistan's Punjab province.

The data for this analysis was obtained from the office of the hydrological directorate located in Lahore, which is responsible for measuring water flows at different barrages in the Punjab province. The data spans 50 to 52 years for two gauging stations located in Pakistani territory. Table 1 provides important characteristics (such as time period of data series, elevation from sea level, and geographic coordinates) of gauging sites considered in the current work. Fig. 2 provides time series plots of the maximum annual flow at each gauging station for a visual representation.

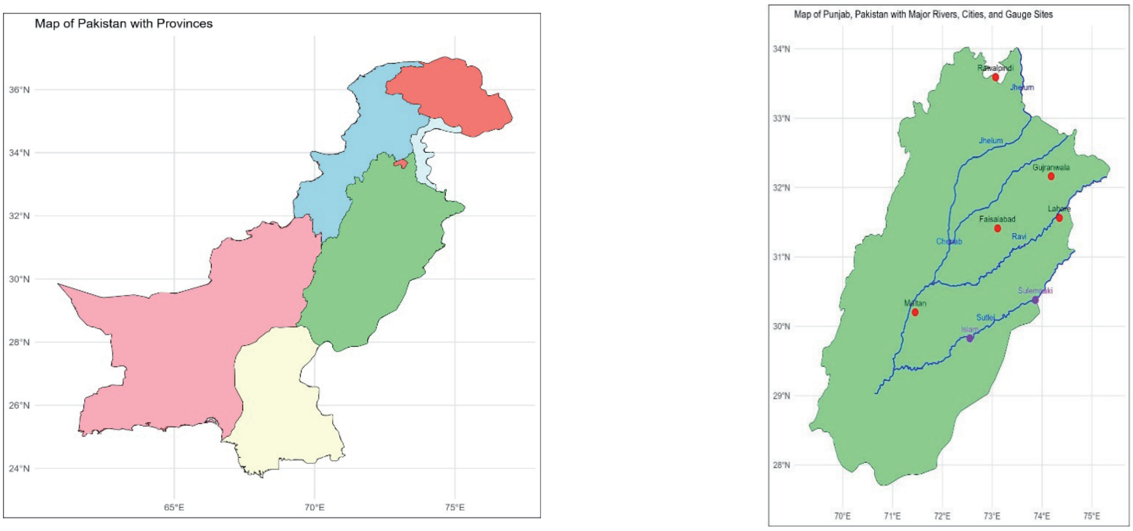

Fig. 1. Map of the Sutlej River with two gauging sites located in Pakistani territory.

Table 1. Basic characteristics of gauging locations.

Gauging location	Time period	Longitude	Latitude	Elevation
Islam	1966-2017	72.5501	29.8258	139 m
Sulemnaki	1966-2017	73.8642	30.3790	160 m

Materials and Methods

An optimal combination of probability distribution and estimation method is crucial for conducting any FFA [2, 38, 39]. This is because the right choice of this blend is vital for efficient modeling and hence accurate estimation of flood levels for different future time periods. We utilized GLO, GUM, REV-GUM, GP, and GEV distributions in our at-site FFA of gauging sites along the Sutlej River. We applied LM, ML, LS, WLS, and RLS methods to estimate the parameters of these distributions. The following subsections briefly discuss the probability distributions and parameter estimation methods being compared.

Candidate Probability Distributions

Several probability distributions, each with its own merits and demerits, have been found useful for FFA across different countries. Table 2 gives the probability density functions $f(x)$, cumulative distribution function (CDF) $F(x)$, and quantile functions $\phi(F)$ of the distributions applied in the current work. These distributions have been applied and found suitable for FFA for different river sites in many studies.

Here, location, scale, and shape parameters are denoted by μ, α, and β, respectively.

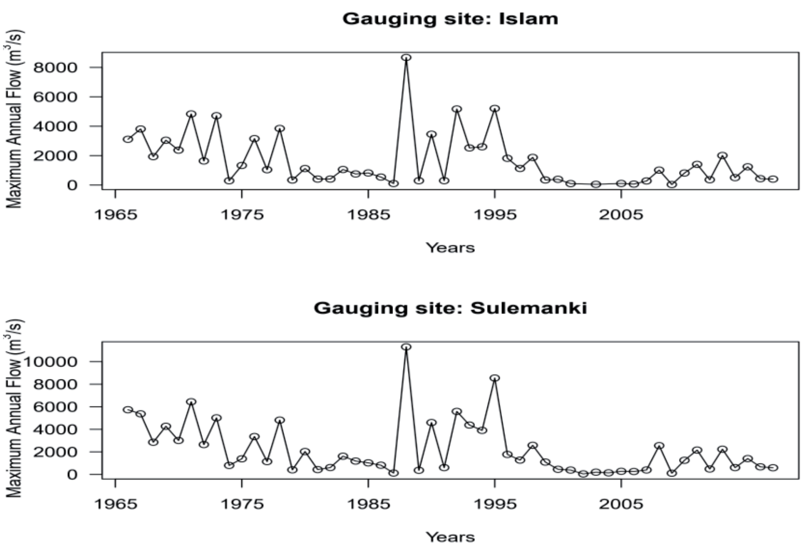

Fig. 2. Time series plots of annual extreme flows at two gauging sites.

Table 2. Probability density, cumulative distribution, and quantile functions of the probability distributions.

Distributions	Probability density function $f(x)$	Cumulative distribution function $F(x)$	Quantile function $\phi(F)$
GLO	$\frac{1}{\alpha}\left[1-\beta\left(\frac{x-\mu}{\alpha}\right)\right]^{(1/\beta-1)}\left[1+\left\{1-\beta\left(\frac{x-\mu}{\alpha}\right)\right\}^{1/\beta}\right]^{-2}$	$\frac{1}{1+\left\{1-\frac{\beta}{\alpha}(x-\mu)\right\}^{\frac{1}{\beta}}}$	$\mu+\frac{\alpha}{\beta}\left[1-\{(1-F)/F\}^{\beta}\right]$
GEV	$\frac{1}{\alpha}\left[1-\beta\left(\frac{x-\mu}{\alpha}\right)\right]^{1/\beta-1}\exp\left\{-\left[1-\beta\left(\frac{x-\mu}{\alpha}\right)\right]^{1/\beta}\right\}$	$\exp\left[-\left\{1-\frac{\beta}{\alpha}(x-\mu)\right\}^{\frac{1}{\beta}}\right]$	$\mu+\frac{\alpha}{\beta}\left[1-(-\log(F))^{\beta}\right]$
GUM	$\frac{1}{\alpha}\exp\left[-\frac{(x-\mu)}{\alpha}-\exp\left(-\frac{(x-\mu)}{\alpha}\right)\right]$	$\exp\left[-\exp\left\{\frac{-(x-\mu)}{\alpha}\right\}\right]$	$\mu-\alpha\{\log(-\log(F))\}$
GPD	$\begin{cases}\frac{1}{\alpha}\left[1-\beta\left(\frac{x-\mu}{\alpha}\right)\right]^{\frac{1}{\beta}-1}, & for\ \beta\neq 0\\ \frac{1}{\alpha}\exp\left(-\frac{(x-\mu)}{\alpha}\right), & for\quad \beta=0\end{cases}$	$1-\left\{1-\frac{\beta}{\alpha}(x-\mu)\right\}^{\frac{1}{\beta}}$	$\mu+\frac{\alpha}{k}\left[1-(1-F)^{\beta}\right]$
REV.GUM	$\frac{1}{\alpha}\exp\left[\frac{(x-\mu)}{\alpha}\right]\exp\left[-\exp\left(\frac{(x-\mu)}{\alpha}\right)\right]$	$1-\exp\left[-\exp\left\{\frac{(x-\mu)}{\alpha}\right\}\right]$	$\mu+\alpha\{\log(-\log(1-F))\}$

where location, scale and shape parameters are denoted by μ, α, and β, respectively.

L-moments Estimation Approach

Pioneered by Hosking [40, 41], the L-moments are computed as direct linear functions of probability-weighted moments (PWM). Although it is a common and easy approach to compute L-moments using PWM [9, 10, 42], L-moments are generally more efficient and convenient compared to PWMs while dealing with practical problems, as they directly quantify the measures of scale and shape of any distribution. For any distribution with quantile function $\phi(F)$, the r[th] PWM B_r is theoretically defined as:

$$\beta_r = \int_0^1 \phi(\mathrm{F})\ \mathrm{F^r dF} \qquad \mathrm{r}=0, 1, 2,... \tag{1}$$

For a given set of n sample values $x_1, x_2, x_3, ..., x_n$, the first four PWMs are calculated as:

$$\beta_0 = n^{-1}\sum_{j=1}^{n} x_{(j)} \tag{2}$$

$$\beta_1 = n^{-1}\sum_{j=2}^{n} \frac{(j-1)}{(n-1)} x_{(j)} \tag{3}$$

$$\beta_2 = n^{-1}\sum_{j=3}^{n} \frac{(j-1)(j-2)}{(n-1)(n-2)} x_{(j)} \tag{4}$$

$$\beta_3 = n^{-1}\sum_{j=4}^{n} \frac{(j-1)(j-2)(j-3)}{(n-1)(n-2)(n-3)} x_{(j)} \tag{5}$$

where $x_{(j)}$ is the j[th] order statistic in the sample of n observations.

Subsequently, the corresponding L-moments are calculated directly through the following relationships:

$$\lambda_1 = \beta_0 \tag{6}$$

$$\lambda_2 = 2\beta_1 - \beta_0 \tag{7}$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \tag{8}$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \tag{9}$$

As a function of L-moments, Hosking [41] proposed L-coefficient of variation (τ_2), L-coefficient of skewness (τ_3), and L-coefficient of kurtosis (τ_4)as:

$$\tau_2 = \frac{\lambda_2}{\lambda_1},\ \tau_3 = \frac{\lambda_3}{\lambda_2},\ \tau_4 = \frac{\lambda_4}{\lambda_2}.$$

Similar to the method of moments estimation, parameters by the L-moments (LM) technique are estimated in a simple and straightforward way. The solution of the system of equations obtained by matching theoretical L-moments with corresponding sample L-moments yields the estimates for unknown parameters.

Maximum Likelihood Estimation Approach

Maximum Likelihood (ML) is one of the most common methods for estimating the parameters of probability distributions. The principle of the ML approach lies in identifying values that maximize the likelihood or log-likelihood function. Consider a random sample of n observations $x_1, x_2, ..., x_n$ from a probability distribution having a density function $f(x_j|\theta)$ with θ being the vector of unknown parameters. The ML estimate of θ is obtained by setting the derivative of the log-likelihood function $\iota(\theta)$ with respect to the elements of θ equal to zero and solving the resultant system of equations. If a closed-form solution is not achievable, numerical optimization techniques can be used to obtain the ML estimates. This study employed the SANN (based on simulated annealing) algorithm for numerical optimization using R-language [43].

Least Squares Estimation Approach

The Least Squares (LS) approach is frequently employed for estimating the parameters of models [44-46]. In the context of probability distributions, the LS approach minimizes the sum of squared difference between corresponding theoretical and observed CDFs. Algebraically,

$$LS(\theta) = \sum_{j=1}^{n}\left[F(x_j) - F_n(x_j)\right]^2, \tag{10}$$

where $F(x_j)$ and $F_n(x_j) = \frac{j}{(n+1)}$ are theoretical and sample CDFs of the data, respectively. The parameter estimates are found by solving the set of simultaneous equations obtained by differentiating this function with respect to unknown parameters. The details of the LS approach are provided in Ali [44]. When complex functions do not result in a closed-form unique solution, optimization algorithms are used to estimate the parameters. We have applied the "optim" function in the R-language for this optimization process.

Weighted Least Squares Estimation Approach

Pioneered by Bergman [47], the Weighted Least Squares (WLS) approach is a modified version of the ordinary LS approach. Unlike the LS method, which applies the same weights to all observations, WLS assigns diverse weights to different observations.

These weights are computed through a certain weight function. In the WLS approach, parameter estimation is done by minimizing the sum of squared weighted difference between theoretical and sample CDFs algebraically,

$$WLS(\theta) = \sum_{j=1}^{n} W_j \left[F(x_j) - F_n(x_j) \right]^2 \tag{11}$$

Following the findings and recommendations of different previous studies [48-50], we have used the following function to compute weights to be assigned to different observations:

$$W_j = \frac{(n+1)^2 (n+2)}{j(n-j+1)} \tag{12}$$

Relative Least Squares Estimation Approach

The Relative Least Squares (RLS) method, introduced by Pablo and Bruce [51], is another modified form of the simple or ordinary LS estimation approach. Recent findings suggest that RLS performs better for different probability distributions [37, 52, 53]. Unlike the LS approach (which minimizes the sum of squared difference between theoretical and sample CDFs), the RLS approach aims to minimize the sum of squared relative difference between the two CDFs. Simply put, through the RLS method, parameters are estimated by minimizing the objective function given as:

$$RLS(\theta) = \sum_{j=1}^{n} \left[\frac{F(x_j) - F_n(x_j)}{F_n(x_j)} \right]^2 \tag{13}$$

Just like in the case of LS and WLS approaches, numerical optimization techniques can be used in case of complex functions that do not result in unique closed-form solutions.

Performance Metrics

The comparative performance of various potential models is evaluated to recommend the most optimal model (blend of distribution and estimation strategy) for a particular gauging location. The comparison among different competing models is done using commonly applied accuracy metrics (like RMSE, RMSPE, MAE, and MAPE) and p-values associated with some goodness-of-fit (GOF) tests (like KS, AD, and CVM). The test statistics for the GOF tests employed are given as:

$$KS = Max\left\{ \left| \hat{F}(x_j) - F_n(x_j) \right| \right\} \tag{14}$$

$$CVM = \frac{1}{12n} + \sum_{j=1}^{n} \left(\frac{2j-1}{2n} - \hat{F}(x_j) \right)^2 \tag{15}$$

$$AD = -n - \sum_{j=1}^{n} \left(\frac{2j-1}{n} \begin{Bmatrix} \log\left(1 - \hat{F}(x_{n-j+1})\right) + \\ \log\left(\hat{F}(x_j)\right) \end{Bmatrix} \right) \tag{16}$$

Similarly, accuracy measures are defined as:

$$RMSE = \sqrt{\frac{\sum_{j=1}^{n} \left(F_n(x_j) - \hat{F}(x_j) \right)^2}{n}} \tag{17}$$

$$RMSE = \frac{100}{\sqrt{n}} \sqrt{\sum_{j=1}^{n} \left(\frac{F_n(x_j) - \hat{F}(x_j)}{F_n(x_j)} \right)^2} \tag{18}$$

$$MAE = \frac{1}{n} \sum_{j=1}^{n} \left| F_n(x_j) - \hat{F}(x_j) \right| \tag{19}$$

$$MAPE = \frac{100}{n} \sum_{j=1}^{n} \left| \frac{F_n(x_j) - \hat{F}(x_j)}{F_n(x_j)} \right| \tag{20}$$

where $F_n(x_j)$ and $\hat{F}(x_j)$ denote the observed and expected CDF of the distribution, respectively. The model with the highest p-value (in the case of the GOF test) and the lowest value (in the case of accuracy measures) is deemed best suited.

After ranking each model based on the above seven performance metrics, a total rank for each model is computed by summing the ranks from these performance indicators as follows:

$$Total\ Rank = \begin{Bmatrix} Rank_{RMSE} + Rank_{RMSPE} + Rank_{MAE} + Rank_{MAPE} \\ + Rank_{KS} + Rank_{AD} + Rank_{CVM} \end{Bmatrix} \tag{21}$$

Finally, for a given gauging location, the model with the highest total rank score is considered the most suitable and optimal model for that particular location.

Estimation of Flood Quantiles

Estimating the maximum flow for a given return period (T) is another vital objective of any FFA. The return period T is generally given in years and indicates how often the peak flow is expected to recur. The estimated flood level x_T corresponding to a return period of T years may be defined as the flood of x_T magnitude and is expected to exceed once in T years. Thus, the cumulative probability of non-exceedance (i.e., the likelihood that the flood level will not be surpassed in a given year) is:

$$F = F(x_T) = P(X \leq x_T) = 1 - P(X \geq x_T) = 1 - 1/T \tag{22}$$

Now, the value F can be inserted in the quantile function of the most suitable model to estimate flood level x_T for return period T. The expressions of quantile functions of different distributions applied in the current study are given in Table 2. The parametric bootstrapping approach [9, 10, 54] is applied to estimate the standard errors of predicted flood quantiles (x_T). Based on these standard errors, 95% confidence intervals are also constructed. The parametric bootstrap approach creates a sampling distribution by generating repeated samples from a particular distribution with estimated parameters. This sampling distribution is then used to compute standard errors for the associated parameter estimates.

Results and Discussion

Table 3 presents the summary statistics for annual maximum flow data for gauging sites considered in the current analysis. For any FFA using probability models, the main assumptions about data series are randomness, independence, stationarity, and skewness [10, 17, 55]. In this study, we have used Wald-Wolfowitz (WW), Augmented Dickey-Fuller (ADF), and autocorrelation coefficient (AC) tests to test the validity of randomness, stationarity, and both independence and skewness, respectively.

The outcomes given in Table 4 clearly indicate that these tests confirm that the data series from both gauging sites satisfy the prerequisite assumptions for conducting at-site FFA.

The parameters estimated (with parametric bootstrap standard errors in parentheses) using various estimation methods for candidate distributions are provided in Tables 5 and 6 for gauging sites located at Islam and Sulemanki, respectively. For both gauging sites, each probability distribution and estimation approach combination was ranked based on accuracy measures and p-values associated with GOF tests. The highest p-value (for GOF tests) and the lowest value (for accuracy measures) were given the highest rank. The total rank score, determined by summing all ranks, is used as a single performance measure for identifying the most plausible model for each gauging site.

The results given in Tables 5 and 6 indicate that the generalized Pareto distribution estimated through the WLS method is the most optimal model for both gauging locations considered for the analysis. Moreover, the same probability distribution estimated with the LS and LM methods is found to be the second-best choice for the Islam and Sulemanki gauging sites, respectively. This outcome contrasts with previous research focused on at-site FFA [9, 10, 56] that identified and recommended different other probability distributions as most suitable for various stations of different rivers analyzed therein.

After identifying the most plausible model for a certain gauging location, the next important step in any at-site FFA is estimating flood levels for various return years. These flood quantiles are determined using the selected model's quantile function and estimated parameters. Table 7 presents the quantile estimates for both gauging locations with certain non-exceedance probabilities (F) for return periods ranging from 5 to 1000 years. The table also provides the standard errors computed through parametric bootstrapping and the resulting 95% confidence intervals associated with the estimated yearly maximum flow. The magnitude of the standard errors for different return years suggests that flood estimates for longer return periods have less precision than those for relatively shorter return periods.

Table 3. Descriptive statistics of the Sutlej River's yearly maximal flow (m^3/s).

Gauging location	n	Mean	Median	Maximum	SD	CV
Islam	50	1665.769	1048.672	8676.99	1791.828	1.075676
Sulemanki	52	2216.793	1260.836	11311.25	2371.395	1.069741

Table 4. Results of testing prerequisite assumptions for data series.

Gauging location	n	ADF (P-value)	WW (P-value)	AC (P-value)	Skewness
Islam	50	0.0437	0.253	0.947	1.6396
Sulemanki	52	0.0373	0.401	0.532	1.6397

Table 5. Parameter estimates and ranking of models (gauging location: Islam).

Distribution	Method	$\hat{\mu}$	$\hat{\alpha}$	$\hat{k}$	Total Rank
GLO	LM	1146.392 (200.513)	721.963 (108.013)	-0.37 (0.114)	100
	MLE	987.485 (4.905)	810.457 (4.689)	-0.844 (0.024)	101
	LS	972.067 (8.584)	804.496 (8.045)	-0.763 (0.164)	128
	WLS	959.722 (5.994)	732.371 (5.858)	-0.705 (0.078)	135
	RLS	975.538 (5.476)	863.958 (5.136)	-0.884 (0.025)	98
GEV	LM	761.986 (162.157)	929.828 (151.955)	-0.289 (0.141)	113
	MLE	747.451 (9.009)	884.708 (8.364)	-0.76 (0.159)	90
	LS	758.278 (10.759)	944.325 (8.73)	-0.441 (0.281)	118
	WLS	758.16 (11.318)	914.308 (8.884)	-0.5 (0.211)	109
	RLS	756.52 (10.333)	951.918 (9.25)	-0.882 (0.191)	70
GUM	LM	904.697 (210.972)	1318.523 (174.77)	------	74
	MLE	903.315 (10.484)	1316.672 (10.821)	------	75
	LS	897.324 (12.271)	1309.558 (10.439)	------	76
	WLS	913.518 (19.586)	1296.689 (14.369)	------	75
	RLS	909.125 (10.643)	1321.682 (10.788)	------	71
GPD	LM	-89.469 (97.313)	1615.762 (414.625)	-0.08 (0.164)	157
	MLE	-67.9 (7.257)	1615.979 (7.796)	-0.044 (0.116)	157
	LS	-94.196 (11.156)	1609.068 (9.367)	-0.143 (0.219)	158
	WLS	**-84.199 (10.106)**	**1623.939 (6.716)**	**-0.122 (0.181)**	**162**
	RLS	-88.865 (11.217)	1617.468 (9.414)	-0.664 (0.906)	48
REV	LM	2426.841 (196.174)	1318.523 (168.585)	------	30
	MLE	2391.641 (8.654)	1312.138 (9.199)	------	36
	LS	2396.879 (11.453)	1320.816 (10.539)	------	34
	WLS	2425.536 (19.841)	1318.764 (14.058)	------	32
	RLS	2446.161 (11.002)	1314.078 (11.449)	------	28

Bootstrap standard errors are given in parentheses

Table 6. Parameter estimates and ranking of models (gauging location: Sulemanki).

Distribution	Method	$\hat{\mu}$	$\hat{\alpha}$	$\hat{k}$	Total Rank
GLO	LM	1525.317 (253.589)	949.012 (136.736)	-0.373 (0.113)	105
	MLE	1233.304 (5.113)	920.598 (5.297)	-0.763 (0.025)	131
	LS	1274.061 (8.205)	1039.922 (8.733)	-0.794 (0.158)	127
	WLS	1262.658 (5.55)	952.484 (5.7)	-0.72 (0.064)	141
	RLS	1647.883 (5.545)	1573.727 (4.949)	-0.961 (0.027)	49
GEV	LM	1020.908 (209.027)	1219.103 (188.793)	-0.294 (0.14)	115
	MLE	1018.684 (9.38)	1185.839 (8.976)	-0.8 (0.157)	77
	LS	1021.778 (10.884)	1229.476 (8.77)	-0.481 (0.272)	113
	WLS	1025.766 (10.81)	1208.012 (7.581)	-0.492 (0.206)	111
	RLS	1012.46 (10.154)	1224.73 (8.032)	-0.835 (0.187)	71
GUM	LM	1211.678 (262.967)	1741.317 (224.538)	------	75
	MLE	1213.103 (10.918)	1740.864 (10.997)	------	70
	LS	1205.364 (11.615)	1734.667 (10.497)	------	80
	WLS	1213.289 (18.297)	1733.667 (13.58)	------	77
	RLS	1213.847 (11.709)	1739.771 (10.819)	------	68
GPD	LM	-92.366 (117.758)	2108.624 (519.333)	-0.087 (0.156)	151
	MLE	-79.171 (6.993)	2118.006 (7.283)	-0.046 (0.112)	146
	LS	-98.561 (10.306)	2105.663 (9.813)	-0.146 (0.227)	148
	WLS	**-87.504 (9.657)**	**2110.886 (6.103)**	**-0.148 (0.185)**	**157**
	RLS	-80.227 (11.674)	2098.269 (9.62)	-0.333 (0.738)	113
REV	LM	3221.908 (246.736)	1741.317 (218.13)	------	28
	MLE	3191.647 (8.823)	1739.562 (9.628)	------	32
	LS	3189.695 (11.361)	1738.235 (10.604)	------	38
	WLS	3213.075 (17.088)	1757.156 (13.873)	------	24
	RLS	3234.048 (10.185)	1736.22 (11.522)	------	28

Bootstrap standard errors are given in parentheses

Table 7. Predicted flood levels for various return years with 95% bootstrap confidence limits.

Gauging location: (Suited model)	Prediction	Non-exceedance probability (F) for different return years							
		0.8	0.9	0.96	0.98	0.99	0.995	0.998	0.9999
		5 years	10 years	25 years	50 years	100 years	200 years	500 years	1000 years
Islam: (GPD with WLS method)	Upper	3564.655	5997.983	10409.8	14997.98	21028.75	28946.63	43346.54	58263.74
	Fit	**2839.004**	**4347.972**	**6648.844**	**8676.331**	**11011.14**	**13722.23**	**18033.71**	**21992.02**
	Lower	2241.69	3098.164	4115.914	4807.061	5437.762	6009.033	6686.533	7144.907
	S.E	335.5952	737.0093	1611.28	2619.033	4029.408	5974.498	9684.63	13695.1
Sulemanki (GPD with WLS method)	Upper	5223.057	9284.002	17492.46	26903.99	40377.46	59725.97	98773.63	143505.9
	Fit	**3854.936**	**6011.172**	**9486.435**	**12751.29**	**16758.13**	**21751.87**	**30459.39**	**39281.62**
	Lower	2843.239	3865.676	5032.844	5794.503	6465.464	7056.528	7731.263	8171.598
	S.E	624.4631	1423.481	3287.559	5611.689	9135.692	14452.37	25866.8	39836.27

Fit: predicted flood quantiles
Upper: lower limit of 95% confidence intervals
Lower: lower limit of 95% confidence intervals
S.E: Bootstrap standard errors

Conclusions

Modeling of yearly peak flow at two gauging locations of the Sutlej River was conducted using some well-known probability distributions estimated using various estimation techniques. The goal was to identify the most apt combination of distribution and estimation approach for each location. We evaluated the performance of GLO, GUM, REV-GUM, GEV, and GP probability distributions estimated using ML, LM, LS, WLS, and RLS estimation approaches to achieve this goal. The comparison was based on a total rank computed based on the rankings of seven different performance metrics. The model with the maximum total rank was selected separately as the most optimal for both locations.

Based on our results, we recommend the generalized Pareto distribution estimated with the weighted least squares method as most suitable for modeling the peak yearly discharge for both gauging locations of the Sutlej River considered in the study.

The findings of the current analysis of water flow data can be utilized to analyze and predict flood levels, manage water reservoirs, and plan hydraulic structures along the Sutlej River and its surrounding catchment area. The model recommended for analyzed gauging locations may also serve as a valuable potential candidate model for at-site or regional FFA of the Sutlej River or other rivers in nearby regions.

Statements and Declarations

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data Availability Statement

The data supporting the study's findings are provided in the supplementary information file with the manuscript.

References

1. ALAM M.A., EMURA K., FARNHAM C., YUAL J. Best-fit probability distributions and return periods for maximum monthly rainfall in Bangladesh. Climate, **6** (1), 1, **2018**.
2. RIZWAN M., GUO S., XIONG F., YIN J. Evaluation of Various Probability Distributions for Deriving Design Flood Featuring Right-Tail Events in Pakistan. Water, **10** (11), 1, **2018**.
3. BADYALINA B., MOKHTAR N.A., AMALINA N., HASSIM N.H. Flood Frequency Analysis using L-Moment for Segamat River. MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics, **37** (2), 47, **2021**.
4. DEY D.K., YAN J. Extreme value modeling and risk analysis: methods and application,. 2016 [Online] CRC Press 2016.
5. GANAMALA K., KUMAR P.S. A case study on flood frequency analysis. International Journal of Civil Engineering and Technology, **8** (4), 1762, **2017**.
6. WU J., LIN B., SHAO Y. Application of regional L-moments analysis method in precipitation frequency analysis for Taihu Lake Basin. Journal of China Hydrology, **35** (5), 15, **2015**.
7. SHAO Y., WU J., LI M. Frequency analysis of extreme precipitation in Huaihe River Basin based on

hydrometeorological regional L-moments metho. Journal of China Hydrology, **36** (6), 16, **2016**.
8. WU Y., XUE L., LIU Y. Local and regional flood frequency analysis based on hierarchical Bayesian model in Dongting Lake Basin, China. Water Science and Engineering, **12** (4), 253, **2019**.
9. KOUSAR S., KHAN A.R., UL HASSAN M., NOREEN Z., BHATTI S.H. Some best-fit probability distributions for at-site flood frequency analysis of the Ume River. Journal of Flood Risk Management, **13** (3), e12640, **2020**.
10. UL HASSAN M., HAYAT O., NOREEN Z. Selecting the best probability distribution for at-site food frequency analysis; a study of Torne River. SN Applied Sciences, **1** (12), 1, **2019**.
11. OLOGHADIEN I. Selection Selection of Probabilistic Model of Extreme Floods in Benue River Basin, Nigeria. European Journal of Engineering and Technology Resear, **6** (1), 7, **2021**.
12. SAEED S.F., MUSTAFA A.S., AUKIDY M.AL. Assessment of Flood Frequency Using Maximum Flow Records for the Euphrates River, Iraq. IOP Conference Series: Materials Science and Engineering 2021[Online] **2021**.
13. BAR-ZEEV N., KING C., PHIRI T., BEARD J., MVULA H., CRAMPIN A.C. Impact of monovalent rotavirus vaccine on diarrhoea-associated post-neonatal infant mortality in rural communities in Malawi: a population-based birth cohort study. The Lancet Global Health, **6** (9), e1036, **2018**.
14. SHAHID A., AHMAD S.R., AHMAD I. Assessment of severe flood events by the use of recent and long-term annual peak discharge data of Chenab River, Pakistan. Arabian Journal of Geosciences, **16** (6), 368, **2023**.
15. MURTAZA D., ROSHNI T., HIMAYOUN D. The investigation of runoff variations and the flood frequency estimates of the Jhelum River, India. Sustainable Water Resources Management, **8** (3), 60, **2022**.
16. LI M., LI X., AO T. Comparative Study of Regional Frequency Analysis and Traditional At-Site Hydrological Frequency Analysis. Water **11** (3), 486, **2019**.
17. BHATTI S.H., UMAR M., SHONGWE S.C., IFRA N., HASSAN M.U. Searching for the most suited distribution and estimation method for at-site flood frequency analysis: A case of the Chenab River. Polish Journal of Environmental Studies, **34** (1), 29, **2025**.
18. LEŠČEŠEN I., UROŠEV M., DOLINAJ D., PABTELIĆ M., TELBISZ T., VERG G., SAVIĆ S., MILOŠEVIĆ D. Regional Flood Frequency Analysis Based on L-Moment Approach (Case Study Tisza River Basin). Water Resources, **46** (6), 853, **2019**.
19. MALEKINEZHAD H., ZARE-GARIZI A. Regional frequency analysis of daily rainfall extremes using L-moments approach. Atmósfera, **27** (4), 411, **2014**.
20. WAGH V., VAIDYA P., AKHARE V., SUREKA H. BOKIL S. Probability distribution of flood frequency analysis. International Research Journal of Engineering and Technology, **7** (8), 3796, **2020**.
21. DRISSIA T.K., JOTHIPRAKASH V., ANITHA A.B. Flood Frequency Analysis Using L Moments: a Comparison between At-Site and Regional Approach. Water Resources Management, **33** (3), 1013, **2019**.
22. XIONG F., GUO S., CHEN L., YIN J., LIU P. Flood frequency analysis using Halphen distribution and maximum entropy. Journal of Hydrologic Engineering, **23** (5), 04018012, **2018**.
23. FAROOQ M., SHAFIQUE M., KHATTAK M.S. Flood frequency analysis of river swat using Log Pearson type 3, Generalized Extreme Value, Normal, and Gumbel Max distribution methods. Arabian Journal of Geosciences, **11** (9), 1, **2018**.
24. KITE G.W. Frequency and Risk Analyses in Hydrology,. 2019[Online] Revised Ed Water Resources Publications, LLC **2019**.
25. CASTELLARIN A., KOHNOVÁ S., GAÁL L., FLEIG A., SALINAS J.L., TOUMAZIS A., KJELDSEN T.R. MACDONALD N. Review of applied statistical methods for flood frequency analysis in Europe, Milestone Report for WG2 of COST Action ES0901. 2012[Online] Tech. rep., Centre for Ecology & Hydrology, Wallingford, UK, on behalf of COST, available at: http://www. cost. eu/media/publications **2012**.
26. SAMANTARAY S., SAHOO A. Estimation of flood frequency using statistical method: Mahanadi River basin, India. H2 Open Journal, **3** (1), 189, **2020**.
27. SHAH A.I., PAN DAS N. Evaluation of probability distribution methods for flood frequency analysis in the Jhelum Basin of North-Western Himalayas, India. Cleaner Water, **2** , 100044, **2024**.
28. BERTOLA M., VIGLIONE A., LUN D., HALL J., BLOSCHL G. Flood trends in Europe: are changes in small and big floods different? Hydrology and Earth System Sciences, **24** (4), 1805, **2020**.
29. CUNNANE C. Statistical distributions for flood frequency analysis. Operational Hydrology Report (WMO), WMO **1989**.
30. BOBEE B., CAVADIAS G., ASHKAR F., BERNIER J., RASMUSSN R. Towards a systematic approach to comparing distributions used in flood frequency analysis. Journal of Hydrology, **142** (1-4), 121, **1993**.
31. KJELDSEN T.R., AHN H., PROSDOCIMI I. On the use of a four-parameter kappa distribution in regional frequency analysis. Hydrological Sciences Journal, **62** (9), 1354, **2017**.
32. AHMAD I., FAWAD M., AKBAR M., ABBAS A., ZAFAR H. Regional Frequency Analysis of Annual Peak Flows in Pakistan Using Linear Combination of Order Statistics. Polish Journal of Environmental Studies, **25** (6), 2255, **2016**.
33. HU L., NIKOLOPOULOS E.I., MARRA F., ANGNOSTOU E.N. Sensitivity of flood frequency analysis to data record, statistical model, and parameter estimation methods:An evaluation over the contiguous United States. Journal of Flood Risk Management, **13** (1), e12580, **2020**.
34. PEREZ G., MANTILLA R., KRAJEWSKI W.F., WRIGHT D.B. Using Physically Based Synthetic Peak Flowsto Assess Local and Regional FloodFrequency Analysis Methods. Water Resources Research, **55** (11), 8384, **2019**.
35. BHAT M.S., ALAM A., AHMAD B., KOTLIA B.S., FAROOQ H., TALOOR A.K., AHMAD S. Flood frequency analysis of river Jhelum in Kashmir basin. Quaternary International, **507** (1), 288, **2019**.
36. DAS S. Goodness-of-fit tests for generalized normal distribution for use in hydrological frequency analysis. Pure and Applied Geophysics, **175** (10), 3605, **2018**.
37. HUSSAIN S., BHATTI S.H., AHMAD T., SHEHZAD M.A. Parameter estimation of the Pareto distribution using least squares approaches blended with different rank methods and its applications in modeling natural catastrophes. Natural Hazards, **107** (2), 1693, **2021**.

38. MOGES S.A., TAYE M.T. Regional flood frequency curves for remote rural areas of the Nile River Basin: The case of Baro-Akobo drainage basin, Ethiopia. In: Extreme Hydrology and Climate Variability Elsevier pp. 385.
39. BHAGAT N. Flood Frequency Analysis Using Gumbel's Distribution Method: A Case Study of Lower Mahi Basin, India. Journal of Water Resources and Ocean Science, **6** (4), 51, **2017**.
40. HOSKING J.R.M. The theory of probability weighted moments,. 1986[Online] IBM Research Division, TJ Watson Research Center, **1986**.
41. HOSKING J.R.M. L-moments: analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society Series B (Methodological), JSTOR, 105, **1990**.
42. HOSKING J.R.M., WALLIS J.R. Regional frequency analysis: an approach based on L-moments,. 2005[Online] Cambridge University Press, **2005**.
43. TEAM R.C. R: A Language and Environment for Statistical Computing. 2022[Online] R Foundation for Statistical Computing. Vienna, Austria, **2022**.
44. ALI S., ARA J., SHAH I. A comparison of different parameter estimation methods for exponentially modified Gaussian distribution. Afrika Matematika, **33** (2), 1, **2022**.
45. KHAN K.H. Parameter Estimation for the Gompertz Distribution Model Using Least-Squares Method in Conjunction with Simplex and Quasi-Newton Optimization Methods. World Applied Sciences Journal, **35** (4), 552, **2017**.
46. ZEINELDIN R.A., CHESNEAU C., JAMAL F., ELGARHY M. Statistical Properties and Different Methods of Estimation for Type I Half Logistic Inverted Kumaraswamy Distribution. Mathematics, **7** (10), 1, **2019**.
47. BERGMAN B. Estimation of Weibull parameters using a weight function. Journal of materials science letters, Springer, **5** (6), 611, **1986**.
48. DEY S., ALZAATREH A., GHOSH I. Parameter estimation methods for the Weibull-Pareto distribution. Computational and Mathematical Methods, **3** (1), e1053, **2021**.
49. USTA I., CELIK E. Comparison of different estimation methods for the Marshall – Olkin extended Weibull distribution. Journal of Scientific Research and Development, **2** (12), 101, **2015**.
50. ZHAO X., ZHANG Z., CHENG W., ZHANG P. A New Parameter Estimator for the Generalized Pareto Distribution under the Peaks over Threshold Framework. Mathematics, **7** (5), 406, **2019**.
51. PABLO B.S., BRUCE E.R. Model Parameter Estimation Using Least Squares. Water Research, **26** (6), 789, **1992**.
52. RASHID M.Z., AKHTER A.S. Estimation Accuracy of Exponential Distribution Parameters. Pakistan Journal of Statistics and Operation Research, University of the Punjab, College of Statistical & Actuarial Science, **7** (2), 217, **2011**.
53. ZAKA A., AKHTAR A.S. Methods for estimating the parameters of Power function distribution. Pakistan Journal of Statistics and Operations Research, **92** (2), 213, **2014**.
54. MEYLAN P., FAVRE A.-C., MUSY A. Predictive hydrology: A frequency analysis approach,. 2012[Online] CRC Press, **2012**.
55. UL HASSAN M., NOREEN Z., AHMED R. Regional frequency analysis of annual daily rainfall maxima in Skåne, Sweden. International Journal of Climatology, **41** (8), 4307, **2021**.
56. AHMAD I., FAWAD M., MAHMOOD I. At-site flood frequency analysis of annual maximum stream flows in Pakistan using robust estimation methods. Polish Journal of Environmental Studies, **24** (6), 2345, **2015**.