Pol. J. Environ. Stud. Vol. 34, No. 1 (2025), 29-42
DOI: *10.15244/pjoes/185970* ONLINE PUBLICATION DATE: 2024-04-03

Original Research

Searching for the Most Suited Distribution and Estimation Method for At-Site Flood Frequency Analysis: A Case of the Chenab River

Sajjad Haider Bhatti[1*], Muhammad Umar[2], Sandile Charles Shongwe[3], Muhammad Irfan[2], Mahmood Ul Hassan[4]

[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
[4]Institute of Environmental Medicine, Division of Biostatistics, Karolinska Institute, Stockholm, Sweden

Received: 8 October 2023
Accepted: 11 March 2024

Abstract

The article deals with at-site flood frequency analysis for different gauging stations of the Chenab River in Pakistan. The study aimed at recommending the most suitable probability distribution and efficient method of parameter estimation for each gauging site. Generalized extreme value, generalized logistic, Gumbel, generalized Pareto, and reverse Gumbel probability models are fitted to the annual peak flow/discharge. For each gauging site, the parameters of these distributions are estimated through L-moments, maximum likelihood, least squares, weighted least squares, and relative least squares methods. For each site, the probability models with a particular estimation method are ranked on the basis of goodness-of-tests and accuracy measures, and then the most suitable pair of model and estimation method is identified through a total rank. The results indicate that the generalized Pareto distribution is the best fit for Marala, Khanki, Qadirabad, and Punjnad, while the generalized extreme value distribution is the most suited for the Trimmu gauging site. As far as the estimation method is concerned, least squares and weighted least squares methods are more accurate for most of the gauging sites. Finally, for each gauging site, the best-suited probability model is used to estimate the annual peak flow and to construct associated confidence intervals for different return years.

Keywords: generalized extreme value, generalized logistic, generalized pareto, maximum likelihood, reverse Gumbel distribution

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

Introduction

Floods are the most destructive natural catastrophes, leading to losses in human lives, infrastructure, crop yields, and industrial production. It can also seriously affect the environment and the social lives of people in terms of electrical damages and risks, contaminated water, sanitary hazards, problems in drainage, and destruction of roads and other communication networks. A flood is a situation when water overflows from a watercourse (river, lake, or stream channel) and comes to the land, which is usually dry [1-4]. The main reasons behind floods are heavy rains, melting glaciers, breakage in reservoirs used for water storage, and a lack of river channels to effectively convey the excess water [5-7]. Floods are natural phenomena occurring at irregular intervals of time. The flood frequency analysis is used with the objective of predicting the recurrence of flood levels for different return periods. Such prediction of floods is important for certain planning at the government and country level, like the construction of water reservoirs, spillways, bridges, canals, and culverts. It is also important to plan disaster management activities (including protection of human lives, assets, and food security) for any possible hazard in the future. Flood frequency can be carried out with different approaches, like the California method, the Hazen method, and Gumbel's method [1]. The flood analysis can be carried out at the country, regional, or site level. At-site flood frequency analysis is the most appropriate, effective, and simple way to predict flood recurrence at a particular gauging station [5, 6, 8-11].

The choice of the probability model blended with an appropriate estimation strategy plays a vital role in the modeling of flood data. Each distribution has its own advantages and disadvantages in describing flood data. There can be different probability distributions with different parameter estimation methods suitable for modeling maximum water flow data from different gauging sites. In the literature concerning at-site flood frequency analysis, different probability distributions like Generalized Extreme Value (GEV), GEV type-1, GEV type-2, normal, generalized normal, two-parameter lognormal, three-parameter lognormal, two-parameter gamma, three-parameter gamma, Pearson type-III, Gumbel, reverse Gumbel, exponential, generalized logistic, four-parameter Wakeby, five-parameter Wakeby, and generalized Pareto have been applied for flood frequency analysis of different rivers in different countries [1, 5, 6, 12-18]. A brief review can be seen in some studies [19-21].

Following the findings of many studies for different rivers in the world, we have used Generalized Logistic (GLO), Generalized Extreme Value (GEV), Gumbel (GUM), Generalized Pareto (GPD), and Reverse Gumbel (REV.GUM) distributions for at-site modeling of annual maximum flood data at 5 sites of the Chenab River. Some of these probability models were also used by Ahmad et al. [22] for at-site analysis of many gauging sites, including those of the Chenab River, but with relatively smaller sequences of peak flow data.

Several estimation methods are available for parameter estimation of probability models used for flood frequency analysis. In the present study, we have applied L-Moments (LM), Maximum Likelihood (ML), Least Squares (LS), Weighted Least Squares (WLS), and Relative Least Squares (RLS) methods for estimating the parameters of considered probability distributions.

The performance of each probability distribution estimated with a particular estimation method is assessed based on different goodness-of-fit tests and accuracy measures.

The present study is focused on at-site flood frequency analysis of the Chenab River, with the prime objective of identifying the most suitable probability distribution with a parameter estimation method for modeling the annual maximum discharge data of different gauging sites of the Chenab River. After identifying the most appropriate probability distribution and estimation method for each gauging station, the maximum annual peak flow for different return periods with a specific non-exceedance probability has been estimated.

The rest of the article is structured as follows: Section 2 provides a brief description of the Chenab River and the data used in the current work. Section 3 describes the candidate probability models and parameter estimation methods. Section 4 presents and discusses the results obtained by applying different probability distributions with estimation methods at different gauging sites. Finally, conclusions drawn from the study are presented in Section 5.

Data and Gauging Sites

The Chenab River is one of the major rivers in Punjab Province, Pakistan. The river flows from northeast to southeast directions and enters Pakistan's Punjab province from Indian Punjab. It measures 960 Km with an approximate enactment area of 29000 km^2. Its average annual flow is 12.38 MAF [23]. A map of the Chenab River showing the considered gauging sites is reproduced from Magsi and Atif [24] in Fig. 1. The data have been collected from the Hydrological Directorate of Discharge and Flood Zone, Lahore. The range of the data is different for different gauging sites, varying from 48 to 90 years. The longitude, latitude, elevation, and length of the data series for each gauging site are given in Table 1. The maximum annual flow at each gauging site is graphically represented in the form of time plots in Fig. 2.

Materials and Methods

The choice of a probability distribution and estimation method is vital in any at-site flood

Fig. 1. A map showing gauging sites of the Chenab Rive (Source: Magsi and Atif, 2012).

Table 1. Brief of gauging sites.

Gauging site	Time period	Longitude	Latitude	Elevation (meter)
Marala	1946-2017	74.4643	32.6723	234
Khanki	1944-2017	73.9687	32.3988	219
Qadirabad	1970-2017	73.6863	32.3193	205
Trimmu	1928-2017	72.1462	31.1447	150
Punjnad	1948-2017	71.0198	29.3464	1402

frequency analysis. In the following sub-section, a brief introduction of candidate probability distributions and estimation methods used in the current study is provided.

Candidate Probability Distributions

The choice of a suitable probability distribution is vital in any at-site flood frequency analysis. Many probability distributions have been used for flood frequency analysis in different countries. For the analysis of five gauging sites of the Chenab River, we have used Generalized Logistic (GLO), Generalized Extreme Value (GEV), Gumbel (GUM), Generalized Pareto (GPD), and reverse Gumbel (REV.GUM) probability distributions. These distributions are applied and found plausible for at-site flood frequency analysis of different rivers and also for other environmental variables [14, 25-30].

L-Moments (LM) Method

Introduced by Hosking [31, 32], the L-moments are linear combinations of probability-weighted moments (PWM). Being the direct measure of the scale and shape of the data, L-moments are more suitable and convenient than probability-weighted moments. However, it is common practice to compute L-moments using PWM [5, 33, 34]. Theoretically, the r[th] PWM β_r, for a probability distribution with a quantile function $\phi(F)$, is defined as:

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

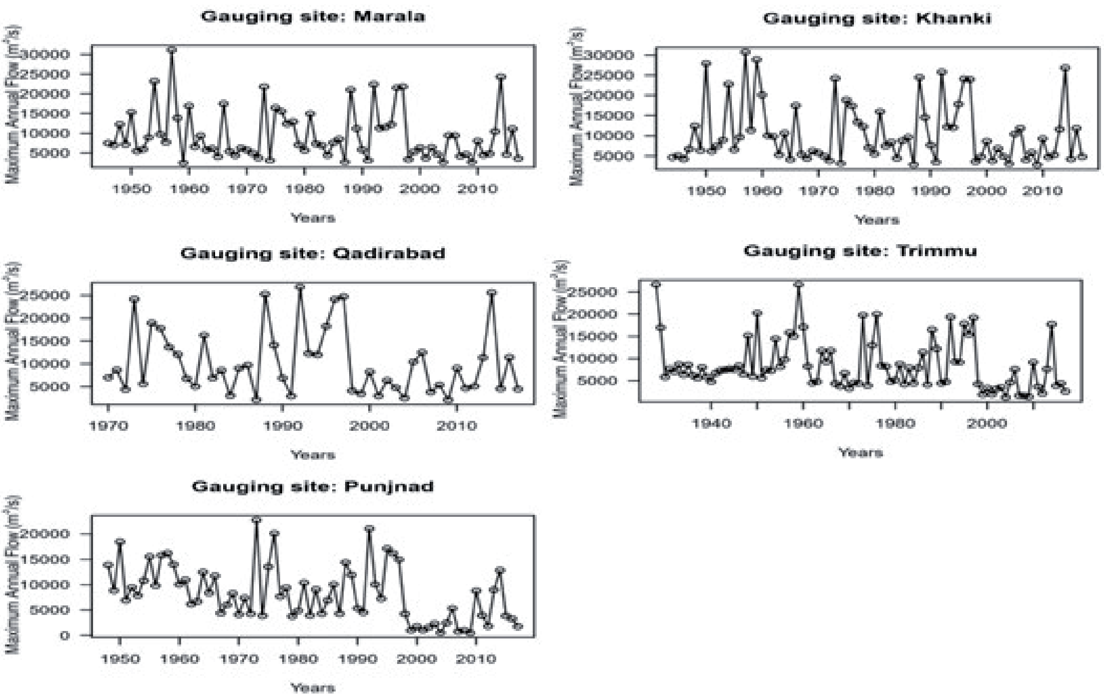

Fig. 2. Time plots of maximum annual discharge at different gauging sites.

For sample data, the first four PWMs are computed as:

$$\beta_0 = n^{-1}\sum_{i=1}^{n} x_{i:n}, \quad \beta_1 = n^{-1}\sum_{i=2}^{n} \frac{(i-1)}{(n-1)} x_{i:n},$$

$$\beta_2 = n^{-1}\sum_{i=3}^{n} \frac{(i-1)(i-2)}{(n-1)(n-2)} x_{i:n},$$

$$\beta_3 = n^{-1}\sum_{i=4}^{n} \frac{(i-1)(i-2)(i-3)}{(n-1)(n-2)(n-3)} x_{i:n}.$$

Consequently, the first four L-moments are computed by the following relationships,

$$\lambda_1 = \beta_0,\ \lambda_2 = 2\beta_1 - \beta_0,\ \lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0,$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0.$$

Hosking [32] defined the coefficients of variation (τ_2), skewness (τ_3) and kurtosis (τ_4) in terms of ratios of the L-moments as:

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

Just like the method of moments, parameter estimation through the L-moments approach is done by solving a system of simultaneous equations obtained by comparing theoretical and corresponding sample L-moments. The LM method has gained much attention recently due to its computational simplicity [5, 6, 12, 35].

Maximum Likelihood (ML) Method

Maximum Likelihood (ML) is a widely applied estimation method. It gives those values as parameter estimates for which the likelihood, or log-likelihood, attains its maximum. Let us have a random sample of n observations $x_1, x_2,\ldots x_n$ from probability density function $f(x_i|\theta)$ where θ is the vector of unknown parameters. Let the likelihood and log-likelihood of n independent observations are $L(\theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$ and $l(\theta) = \sum_{i=1}^{n} \log f(x_i \mid \theta)$, respectively. The ML estimate of θ is the value for which $l(\theta)$ attains its maximum. The maximization is achieved by equating the derivative of $l(\theta)$ with respect to θ to zero. If this process does not yield a closed-form solution, then maximization can be achieved numerically through an optimization technique. For the current work, we have applied algorithms available in R-language [36] for numerical optimization.

Least Squares Method

The Least Squares (LS) method is commonly used for parameter estimation in probability models [37-39]. It is based on minimizing the following objective function:

$$LS(\theta) = \sum_{i=1}^{n}\left[F(x_i) - \frac{i}{n+1}\right]^2,$$

where $F(x_i)$ is the theoretical CDF of the corresponding probability distribution while $i/(n+1)$ is the sample or observed CDF of the data. This can be done by solving the system of equations obtained by taking derivatives of the above function with respect to unknown parameters. The details can be seen in Ali [37]. For complex functions, the parameter estimates can be obtained by a suitable optimization algorithm.

Weighted Least Squares Method

As proposed by Bergman [40], Weighted Least Squares (WLS) is a variant of the LS method. Instead of identical weights for all observations, WLS uses different weights for different observations. Specifically, in the WLS method, parameter estimates are obtained by minimizing the following function:

$$WLS(\theta) = \sum_{i=1}^{n} W_i\left[F(x_i) - \frac{i}{n+1}\right]^2.$$

Following different studies [41-44], we have used weight function as:

$$W_i = \frac{(n+1)^2(n+2)}{i(n-i+1)}.$$

Relative Least Squares Method

Proposed by Pablo and Bruce [45], Relative Least Squares (RLS) is another variant of the LS method. Instead of squared errors (as in the LS method), the RLS method is based on minimizing the squared relative errors. Recently, RLS is reported to be a better performer, in the case of some two-parameter probability distributions [46-48]. Parameter estimates through the RLS method can be obtained by minimizing the following function:

$$RLS(\theta) = \sum_{i=1}^{n}\left[\frac{F(x_i) - \frac{i}{n+1}}{\frac{i}{n+1}}\right]^2.$$

Performance Indicators

To suggest the most suitable probability model for a particular gauging site, the performance of candidate models with an estimation method is assessed. The performance comparison is carried out using some common measures of accuracy and goodness-of-fit (GOF) tests. The test statistics of the GOF tests used in the current study are defined as:

Kolmogorov-Smirnov (KS) test,

$$KS = Max\left\{\left|F_n(x_i) - \hat{F}(x_i)\right|\right\}.$$

Anderson-Darling (AD) test,

$$CVM = \frac{1}{12n} + \sum_{i}^{n}\left(\frac{2i-1}{2n} - \hat{F}(x_i)\right)^2.$$

Cramér–von Mises (CVM),

$$AD = -n - \sum_{i=1}^{n}\left(\frac{2i-1}{n}\left\{\log\left(1-\hat{F}(x_{n-i+1})\right) + \log\left(\hat{F}(x_i)\right)\right\}\right).$$

Similarly, accuracy measures like Root Mean Square Error (RMSE), Root Mean Square Percentage Error (RMSPE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) are defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{n}}, \quad RMSPE = \sqrt{\sum_{i=1}^{n}\left(\frac{x_i - \hat{x}_i}{x_i}\right)^2} * \frac{100}{\sqrt{n}},$$

$$MAE = \frac{100}{n}\sum_{i=1}^{n}\left|x_i - \hat{x}_i\right|, \quad MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{x_i - \hat{x}_i}{x_i}\right|.$$

In the above GOF tests and accuracy measures, $F_n(x_i)$ represents the observed or empirical CDF and $\hat{F}(x_i)$ is the expected CDF of the distribution with particular parameter estimates. Similarly, x_i represents the observed value, while $\hat{x}_i$ is the estimated value using particular parameter estimates.

Estimation of Quantiles

The estimation of maximum peak flow for a particular return period (T) x_T is a major objective of any flood frequency analysis. The recurrence interval T is expressed in terms of years, in which the maximum peak flow x_T is expected to return. A specific flood level x_T (T being the return period) may be exceeded once in T years. Therefore, $P(X > x_T) = \frac{1}{T}$, then the cumulative probability of non-exceedance $F(x_T)$ is

$$F = F(x_T) = P(X \leq x_T) = 1 - P(X > x_T) = 1 - 1/T$$

The quantile x_T can be estimated by putting F in the quantile function of the corresponding probability distribution. If the quantile function of a probability distribution cannot be expressed in closed form, then the numerical relationships between x_T and F can be used to evaluate x_T for a given value of F. The quantile estimate for a return period of T years is just computed by inserting value $F = (1–1/T)$ in the quantile function. The standard errors of these flood levels for different return years were estimated, and hence 95% confidence intervals were constructed using the parametric bootstrapping approach [5, 6, 49].

Results and Discussion

The descriptive statistics of maximum annual discharge data from different gauging stations are given in Table 2. From these summary measures, it can be seen that, on average, Khanki gauging station has the maximum annual peak flow, and this particular site also has a higher variation in the maximum yearly flow. Generally, we see that variability in annual maximum peak flow is not too different for different gauging stations, which is evident from values of the coefficient of variation ranging from 66.45% to 70.25%. Randomness, independence, stationarity, and skewness are the core assumptions for the data to be used in flood frequency analysis through probability models [6, 49, 50]. In the current article, we used the Wald-Wolfowitz (WW) test for testing randomness and the correlation coefficient at lag-1 for assessing independence, while the assumption of stationarity was tested using the Augmented Dickey-Fuller (ADF) test. The results of testing these assumptions are given in Table 3. From the p-values of the ADF and WW tests in Table 3, it is clear that assumptions of stationarity and randomness seem reasonable for all gauging sites considered in the study. Moreover, values of serial correlation at lag-1 do indicate that there is no serious/severe violation of the independence assumption. Similarly, values of the coefficient of skewness indicate that the data series for all gauging stations is positively skewed.

The validity of these assumptions justified the application of considered probability distributions for modeling of the annual maximum peak flow data for all gauging sites.

The parameter estimates through LM, ML, LS, WLS, and RLS methods for all of the candidate probability distributions are given in Tables 4-8 for gauging sites at Marala, Khanki, Trimmu, Qadirabad, and Punjnad, respectively. The parameter estimates are accompanied by standard errors (in parenthesis) computed using the parametric bootstrapping approach [5, 6, 49]. The comparative performance of different probability distributions with estimation methods is also given in Tables 4-8. The performance is assessed in terms of the total number of ranks based on three GOF tests and four accuracy measures. For each gauging site, the blend of

Table 2. Descriptive statistics of annual maximum discharge (m^3/s) of the Chenab River.

Gauging site	n	Mean	Median	Maximum (Annual)	SD	CV
Marala	72	9418.87	7080.29	31148.53	6315.72	0.6705
Khanki	74	10569.08	8004.85	30765.12	7424.82	0.7025
Qadirabad	48	10231.01	8417.45	26859.10	7170.63	0.7009
Trimmu	90	8452.82	7326.77	26731.10	5671.74	0.6709
Punjnad	70	8208.12	7551.86	22724.72	5454.29	0.6645

Table 3. Results of testing assumptions of data from different gauging sites.

Gauging site	n	ADF(P-value)	WW(P-value)	Correlation at lag-1	Skewness
Marala	72	<0.01	0.2353	0.001	1.2856
Khanki	74	<0.01	0.8149	0.047	1.1508
Qadirabad	48	<0.01	1.00	0.079	0.9622
Trimmu	90	<0.01	0.3964	0.287	1.2185
Punjnad	70	<0.01	0.0541	0.295	0.6158

probability model and estimation method is compared by ranking based on GOF tests and accuracy measures. The highest rank is given to the highest p-value (in the case of GOF tests) and the lowest value (in the case of accuracy measures). The total rank score computed by adding all ranks is given for different combinations of probability distribution and estimation methods. The probability distribution with the maximum total rank is considered to be the most suitable for the particular gauging site.

From these tables, it turns out that Generalized Pareto (GP) distribution with LS estimation is more suitable for Marala and Khanki gauging stations, as this blend of probability distribution and estimation method got the highest rank based on seven different performance indicators. For gauging sites at Qadirabad and Punjnad, GP distribution with the WLS method provides more accurate and precise results. Generalized extreme value distribution with ML estimation is best

Table 4. Parameter estimates and performance comparison (Gauging site: Marala).

Distribution	Method	$\hat{\mu}$	$\hat{\alpha}$	$\hat{k}$	Total Rank
GLO	LM	7767.663 (612.488)	2811.449 (331.925)	-0.316 (0.092)	102
	MLE	7685.134 (5.892)	3105.981 (5.67)	-0.55 (0.035)	83
	LS	7298.628 (8.7)	3038.087 (8.322)	-0.561 (0.138)	113
	WLS	7324.065 (5.009)	2902.162 (4.628)	-0.503 (0.076)	136
	RLS	7355.782 (4.984)	3140.995 (4.89)	-0.582 (0.034)	96
GEV	LM	6230.542 (529.408)	3770.856 (479.633)	-0.216 (0.111)	126
	MLE	6231.38 (9.024)	3769.103 (8.699)	-0.398 (0.12)	107
	LS	6210.904 (10.72)	3787.61 (9.787)	-0.348 (0.231)	120
	WLS	6222.21 (7.434)	3777.271 (6.179)	-0.33 (0.178)	122
	RLS	6230.72 (9.075)	3773.934 (7.98)	-0.488 (0.188)	72

Table 4. Continued.

GUM	LM	6642.637 (616.005)	4809.696 (504.781)	---	73
	MLE	6645.577 (10.808)	4806.611 (11.215)	---	70
	LS	6629.655 (12.357)	4803.194 (10.47)	---	79
	WLS	6643.661 (14.906)	4791.505 (12.161)	---	76
	RLS	6652.114 (11.902)	4809.471 (11.106)	---	67
GPD	LM	2622.268 (296.351)	7059.481 (1346.109)	0.039 (0.137)	142.5
	MLE	2637.917 (6.597)	7054.279 (7.41)	0.067 (0.086)	136.5
	LS	**2608.045 (11.247)**	**7061.501 (10.225)**	**0.014 (0.173)**	**146.5**
	WLS	2614.224 (6.396)	7062.709 (5.011)	0.013 (0.129)	144.5
	RLS	2619.634 (10.667)	7061.501 (9.184)	-0.059 (1.293)	125.5
REV	LM	12195.101 (641.604)	4809.696 (511.94)	---	22.5
	MLE	12191.336 (8.003)	4802.985 (8.62)	---	28.5
	LS	12190.788 (11.637)	4808.515 (10.064)	---	30.5
	WLS	12190.117 (15.513)	4823.409 (11.786)	---	27.5
	RLS	12194.263 (10.3)	4798.256 (12.029)	---	28.5

Table 5. Parameter estimates and performance comparison (Gauging site: Khanki).

Distribution	Method	$\hat{\mu}$	$\hat{\alpha}$	$\hat{k}$	Total Rank
GLO	LM	8578.07 (667.167)	3297.339 (386.536)	-0.323 (0.093)	96
	MLE	8327.079 (4.872)	3719.802 (4.996)	-0.641 (0.027)	77
	LS	7829.055 (8.756)	3603.425 (8.066)	-0.65 (0.135)	97
	WLS	7824.397 (4.275)	3305.906 (4.286)	-0.569 (0.069)	116
	RLS	7804.565 (4.714)	3520.225 (4.621)	-0.649 (0.029)	110
GEV	LM	6781.488 (596.143)	4398.43 (528.019)	-0.226 (0.11)	115
	MLE	6779.137 (8.82)	4387.121 (8.658)	-0.554 (0.115)	79
	LS	6770.721 (10.727)	4406.267 (9.219)	-0.37 (0.247)	114
	WLS	6782.714 (7.262)	4397.79 (6.158)	-0.355 (0.175)	115
	RLS	6760.904 (9.421)	4379.058 (7.609)	-0.634 (0.17)	59

Table 5. Continued.

GUM	LM	7287.04 (708.425)	5685.983 (611.485)	---	73
	MLE	7286.831 (11.241)	5671.209 (11.493)	---	81
	LS	7281.227 (12.665)	5693.248 (11.419)	---	77
	WLS	7287.752 (14.762)	5669.175 (12.005)	---	77
	RLS	7289.501 (11.638)	5685.56 (11.303)	---	67
GPD	LM	2597.849 (372.489)	8150.795 (1491.281)	0.022 (0.125)	164
	MLE	2622.528 (6.373)	8140.975 (7.507)	0.065 (0.08)	140
	LS	**2606.5 (10.996)**	**8156.522 (10.048)**	**0.005 (0.17)**	**170**
	WLS	2597.743 (6.112)	8152.081 (4.5)	-0.001 (0.133)	160
	RLS	2598.673 (11.097)	8154.775 (8.818)	-0.012 (0.502)	153
REV	LM	13851.117 (653.604)	5685.983 (577.742)	---	24
	MLE	13847.455 (8.518)	5680.258 (9.077)	---	28
	LS	13840.106 (12.046)	5692.337 (9.851)	---	27
	WLS	13812.404 (14.578)	5687.364 (12.066)	---	35
	RLS	13855.784 (11.213)	5682.703 (12.367)	---	21

Table 6. Parameter estimates and performance comparison (Gauging site: Qadirabad).

Distribution	Method	$\hat{\mu}$	$\hat{\alpha}$	$\hat{k}$	Total Rank
GLO	LM	8532.595 (943.523)	3447.037 (482.036)	-0.274 (0.101)	89
	MLE	8140.054 (6.105)	3823.902 (6.241)	-0.592 (0.043)	81
	LS	7916.106 (8.888)	3906.402 (8.014)	-0.544 (0.165)	103
	WLS	7880.047 (5.242)	3581.542 (5.366)	-0.488 (0.102)	119
	RLS	7824.25 (5.756)	4133.924 (5.76)	-0.667 (0.04)	64
GEV	LM	6609.171 (760.99)	4781.896 (676.025)	-0.156 (0.122)	112
	MLE	6617.385 (9.528)	4785.468 (9.1)	-0.372 (0.145)	97
	LS	6602.378 (10.34)	4783.172 (8.941)	-0.283 (0.268)	116
	WLS	6589.003 (7.828)	4786.033 (6.999)	-0.263 (0.22)	126
	RLS	6603.9 (9.475)	4783.234 (8.642)	-0.508 (0.242)	68

Table 6. Continued.

GUM	LM	6973.139 (837.479)	5644.117 (760.89)	---	70
	MLE	6960.049 (11.416)	5634.125 (11.208)	---	78
	LS	6954.55 (11.025)	5636.569 (10.718)	---	79
	WLS	6984.612 (13.439)	5635.43 (11.65)	---	70
	RLS	6972.529 (11.505)	5632.61 (11.148)	---	78
GPD	LM	1858.87 (518.891)	9544.299 (2122.91)	0.14 (0.154)	162
	MLE	1866.996 (6.332)	9536.78 (7.265)	0.206 (0.084)	142
	LS	1841.839 (10.652)	9522.373 (9.718)	0.104 (0.179)	160
	WLS	**1864.111 (6.614)**	**9538.813 (5.112)**	**0.122 (0.141)**	**171**
	RLS	1860.427 (11.3)	9535.66 (9.081)	0.112 (1.109)	165
REV	LM	13488.884 (876.851)	5644.117 (709.547)	---	20
	MLE	13488.048 (8.474)	5643.668 (9.037)	---	25
	LS	13489.536 (12.453)	5659.823 (9.925)	---	26
	WLS	13468.959 (13.579)	5629.302 (11.608)	---	32
	RLS	13508.622 (10.374)	5615.216 (12.053)	---	22

Table 7. Parameter estimates and performance comparison (Gauging site: Trimmu).

Distribution	Method	$\hat{\mu}$	$\hat{\alpha}$	$\hat{k}$	Total Rank
GLO	LM	7137.086 (520.59)	2647.512 (267.601)	-0.276 (0.075)	112
	MLE	6884.571 (6.052)	2668.764 (6.829)	-0.388 (0.041)	129
	LS	6859.404 (9.089)	2769.048 (8.109)	-0.419 (0.123)	118
	WLS	6903.337 (4.801)	2732.891 (4.616)	-0.375 (0.072)	137
	RLS	6985.346 (5.814)	2810.131 (5.347)	-0.39 (0.046)	125
GEV	LM	5661.177 (455.562)	3666.866 (382.674)	-0.158 (0.093)	142
	MLE	**5658.47 (8.095)**	**3650.411 (8.223)**	**-0.205 (0.09)**	**149**
	LS	5646.884 (11.211)	3656.956 (8.9)	-0.232 (0.2)	129
	WLS	5656.272 (7.13)	3666.872 (5.656)	-0.217 (0.143)	136
	RLS	5650.45 (9.273)	3657.876 (8.406)	-0.226 (0.164)	135

Table 7. Continued.

GUM	LM	5945.849 (497.13)	4343.213 (432.186)	---	67
	MLE	5945.236 (11.123)	4335.3 (11.461)	---	73
	LS	5928.813 (12.522)	4354.189 (10.629)	---	76
	WLS	5941.38 (17.229)	4334.401 (13.378)	---	78
	RLS	5958.096 (11.527)	4330.818 (10.811)	---	66
GPD	LM	2025.099 (273.118)	7296.174 (1201.291)	0.135 (0.114)	106.5
	MLE	2027.821 (5.886)	7307.77 (6.828)	0.16 (0.064)	97.5
	LS	2016.544 (11.502)	7286.063 (10.02)	0.128 (0.138)	108.5
	WLS	2017.511 (5.55)	7294.157 (4.441)	0.116 (0.108)	108.5
	RLS	2013.246 (10.849)	7300.421 (9.2)	0.281 (0.678)	64.5
REV	LM	10959.79 (511.152)	4343.213 (408.4)	---	22.5
	MLE	10959.79 (8.378)	4343.213 (9.275)	---	22.5
	LS	10961.479 (11.472)	4346.723 (10.155)	---	18.5
	WLS	10948.152 (17.279)	4340.959 (13.053)	---	30.5
	RLS	10963.743 (10.636)	4330.726 (11.514)	---	23.5

Table 8. Parameter estimates and performance comparison (Gauging site: Punjnad).

Distribution	Method	$\hat{\mu}$	$\hat{\alpha}$	$\hat{k}$	Total Rank
GLO	LM	7468.051 (671.275)	2977.291 (317.703)	-0.147 (0.07)	67
	MLE	7431.049 (6.926)	3144.326 (7.18)	-0.285 (0.058)	75
	LS	7341.052 (8.586)	3466.268 (7.841)	-0.224 (0.133)	90
	WLS	7272.286 (5.094)	3205.575 (4.879)	-0.243 (0.103)	86
	RLS	7164.378 (5.399)	4421.462 (4.846)	-0.6 (0.036)	39
GEV	LM	5711.289 (602.06)	4596.079 (467.391)	0.035 (0.088)	114
	MLE	5714.788 (8.1)	4576.992 (8.463)	0.026 (0.088)	114
	LS	5719.329 (10.749)	4604.797 (8.647)	-0.031 (0.188)	115
	WLS	5706.757 (6.21)	4599.097 (5.75)	-0.013 (0.139)	119
	RLS	5699.688 (10.232)	4591.782 (8.298)	-0.235 (0.219)	67

Table 8. Continued.

GUM	LM	5637.989 (581.505)	4452.644 (487.011)	---	105
	MLE	5657.312 (11.528)	4448.414 (10.605)	---	109
	LS	5625.473 (11.877)	4462.63 (10.601)	---	110
	WLS	5659.205 (15.247)	4466.814 (12.678)	---	114
	RLS	5641.932 (11.274)	4451.137 (10.009)	---	106
GPD	**LM**	**534.538 (424.809)**	**11405.314 (2036.261)**	**0.486 (0.144)**	**143**
	MLE	535.504 (5.419)	11403.13 (5.824)	0.48 (0.046)	141
	LS	542.904 (10.416)	11422.739 (9.598)	0.49 (0.104)	138
	WLS	**533.236 (4.439)**	**11405.137 (3.35)**	**0.479 (0.07)**	**143**
	RLS	530.215 (11.619)	11403.793 (8.518)	0.444 (1.924)	135
REV	LM	10778.26 (574.289)	4452.644 (466.274)	---	27
	MLE	10774.825 (8.435)	4446.246 (8.978)	---	28
	LS	10777.054 (11.633)	4451.614 (10.348)	---	30
	WLS	10751.577 (14.371)	4457.8 (11.693)	---	33
	RLS	10774.558 (9.962)	4440.219 (12.089)	---	27

Table 9. Estimated flood quantiles for different return years with 95% confidence intervals.

		Non-exceedance probability (F) for different return years							
Gauging Site: (Suited Distribution)	Value	0.8	0.9	0.96	0.98	0.99	0.995	0.998	0.999
		5 years	10 years	25 years	50 years	100 years	200 years	500 years	1000 years
Marala (GPD-LS)	Upper	17949.04	27757.82	45021.34	62343.73	84473.74	112745.7	162504.5	212433
	Fit	14220.37	19501.97	26957.69	33098.85	39824.25	47309.51	58762.06	69000.38
	Lower	11789.67	14657.23	17702.1	19548.13	21075.2	22338.43	23679.77	24492.99
	S.E	1647.587	3520.446	7427.179	11814.1	17913.43	26421.17	43318.91	62844.6
Khanki (GPD-LS)	Upper	20667.49	32509.69	53739.87	75428.17	103573.4	140097.8	205577.2	272469.6
	Fit	16128.16	22377.37	31285.05	38689.1	46854.56	55995.15	70040.55	82606.77
	Lower	13412.65	16887.34	20658.8	22992.36	24956.54	26609.81	28400.71	29508.81
	S.E	1919.534	4116.677	8694.007	13784.1	20728.77	30117.11	47727.62	66569.8
Qadirabad (GPD-WLS)	Upper	19765.38	29280.29	43891.51	56717.94	71304.91	87894.08	113368.6	135731.4
	Fit	16104.75	21707.11	28789.4	33979.15	39084.32	44155.13	50887.19	56062.28
	Lower	13417.4	16626.29	19779.4	21545.14	22910.33	23965.85	25003.01	25583.82
	S.E	1648.671	3271.252	6196.686	9003.383	12346.7	16269.48	22453.1	27997.8

Table 9. Continued.

Trimmu (GEV-MLE)	Upper	13054.41	18645.49	28389.66	38217.61	50912.66	67374.96	96896.5	127063
	Fit	12101.51	16177.22	22391.4	27947.25	34442.37	42088.2	54398.57	65788.5
	Lower	11246.2	14105.83	17799.63	20564.26	23344.67	26151.79	29931.2	32851.92
	S.E	462.701	1165.032	2735.456	4580.798	7199.159	10847.86	17859.12	25457.34
Punjnad (GPD-WLS)	Upper	14811.35	18987.19	23236.55	25703.95	27670.58	29238.06	30832.95	31758.61
	Fit	13463.09	16678.74	19634.62	21184.7	22321.27	23156.64	23933.69	24346.23
	Lower	12346.12	14833.74	16882.35	17837.64	18470.24	18889.14	19234.18	19395.3
	S.E	640.8825	1080.529	1653.592	2047.263	2394.861	2694.457	3020.988	3221.168

Fit: estimated flood quantiles;
S.E: bootstrap standard errors;
Lower (Upper): lower (upper) limits of 95% confidence intervals for estimated flood quantiles.

suited for the gauging site at Trimmu. It is noted that the GP distribution with LM and WLS estimations performs equally well for Punjnad.

From these results, there is no single probability distribution that fits best for all gauging sites. This finding is in line with the other studies focused on at-site flood frequency analysis [5, 6, 22]. From the results, it is also clear that GP distribution with LS and WLS estimation is a suitable model for most of the sites. As there were varying sample sizes for different sites, it means that sample size does not have an impact on the performance of GP distribution. It is interesting to note that for all gauging sites, the most suitable probability model with one estimation method is also among the top performers with other estimation methods.

Once the best-fit probability distribution is chosen for any particular site, the next objective of the flood frequency analysis is to compute quantiles for different return periods. Such quantiles for different return periods are computed using quantile functions and parameter estimates of the most appropriate blend of probability distribution and method of estimation. For each gauging site, the quantile estimate x_T with non-exceedance probability F for return periods 5, 10, 25, 50, 100, 200, 500, and 1000 years is given in Table 9. Uncertainties are always associated with the estimated flood quantiles computed for different return periods. To give an idea of these uncertainties, Table 9 also contains the 95% confidence intervals for estimated flood quantiles using parametric bootstrapping standard errors. For all gauging stations, the values of standard errors indicate that there is more uncertainty in flood estimates for longer return periods as compared to estimates of relatively short return periods.

Conclusions

The modeling of maximum annual discharge at different gauging sites of the Chenab River is done through different probability models blended with different estimation methods. To recommend the best-suited pair of probability models and methods of estimation for each gauging site, we compared the performance of GLO, GEV, GUM, GP, and REV. GUM distributions were estimated with the LM, ML, LS, WLS, and RLS methods. The comparison is based on the total rank of 7 different measures of fit. For each gauging site, the probability distribution with the highest total rank score is considered most appropriate.

Different probability distributions emerge as most suitable for flood frequency analysis at different sites. Generalized Pareto with LS estimation is the most suitable distribution for Marala and Khanki gauging sites. For gauging sites at Qadirabad and Punjnad, GP distribution with the WLS method is found to be the most suitable combination, while GEV distribution with the ML method performs better than all other distributions for the Trimmu gauging station. It can be concluded that GPD with LS and WLS estimation is the most suitable combination for most of the gauging stations (in fact, for all sites except for the gauging site at Trimmu). The results are well in line with many studies focused on at-site flood frequency analysis of rivers in different countries. The emergence of different best-suited probability distributions for different gauging sites has been reported in many other studies [5, 6, 9-11, 22]. Similarly, our finding about the larger uncertainty associated with longer return periods is common and reported in different studies [5, 6].

These results can be used to study and forecast flood levels, the management of water reservoirs, and the planning of hydraulic structures at the Chenab River (and related catchment areas). The most suitable distributions recommended for different gauging sites can be useful candidate distributions for regional analysis of the Chenab River or at-site analysis of other rivers in near geographical locations.

Limitations

Like every research project, the present study also has certain limitations, and work can be done in the future to overcome them. For instance, some other probability distributions, particularly four or five parameter distributions, can be applied. Similarly, different other estimation methods can be used for parameter estimation, like the method of maximum product spacing, Bayesian estimation, etc. Moreover, the analysis can be improved by incorporating a multivariate structure by taking temperature, rainfall, humidity, and other environmental factors into account.

Conflict of 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.

Acknowledgments

The authors are grateful to Dr. Shahid Mehmood (University of Oxford) for proof reading the article for English language. This improved the understanding for the reader a lot.

References

1. GANAMALA K., KUMAR P.S. A case study on flood frequency analysis. International Journal of Civil Engineering and Technology, **8** (4), 1762, **2017**.
2. 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**.
3. 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**.
4. 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**.
5. KOUSAR S., KHAN A.R., UL HASSAN M., 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**.
6. 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**.
7. BAIDYA S., SINGH A., PANDA S.N. Flood frequency analysis. Natural Hazards, **100** (3), 1137, **2020**.
8. LI M., LI X., AO T. Comparative Study of Regional Frequency Analysis and Traditional At-Site Hydrological Frequency Analysis. Water, **11** (3), 486, **2019**.
9. BADYALINA B., MOKHTAR N.A., JAN N.A.M., HASSIM N.H., YUSOP H. Flood Frequency Analysis using L-Moment for Segamat River. MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics, **37** (2), 47, **2021**.
10. ZHANG Z., STADNYK T.A., BURN D.H. Identification of a preferred statistical distribution for at-site flood frequency analysis in Canada. Canadian Water Resources Journal, **45** (1), 43, **2020**.
11. 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**.
12. 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**.
13. GURU N., JHA R. Flood Frequency Analysis of Tel Basin of Mahanadi River System, India Using Annual Maximum and POT Flood Data. Aquatic Procedia, Elsevier **4** , 427, **2015**.
14. VIVEKANANDAN N. Flood frequency analysis using method of moments and L-moments of probability distributions. Cogent Engineering, **2** (1), 1018704, **2015**.
15. 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**.
16. 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, **1076** (2021), 1, **2021**.
17. SAMANTARAY S., SAHOO A. Estimation of flood frequency using statistical method: Mahanadi River basin, India. H2 Open Journal, **3** (1), 189, **2020**.
18. SINAM R. At Site Flood Frequency Analysis of Baitarani River at Champua Watershed, Odisha. International Journal of Scientific Research in Science and Technology, **6** (6), 54, **2019**.
19. 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, Centre for Ecology & Hydrology, Wallingford, UK, pp. 142, **2012**.
20. CUNNANE C. Statistical distributions for flood frequency analysis. Operational Hydrology Report, WMO, **1989**.
21. BOBEE B., CAVADIAS G., ASHKAR F., BERNIER J., RASMUSSEN P. Towards a systematic approach to comparing distributions used in flood frequency analysis. Journal of Hydrology, **142** (1-4), 121, **1993**.
22. 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**.
23. RIAZ M., AZIZ A., HUSSAIN S. Flood Forecasting of an Ungauged Trans-boundary Chenab River Basin Using Distributed Hydrological Model Integrated Flood Analysis

System (IFAS). Pakistan Journal of Meteorology, **13** (26), 51, **2017**.
24. MAGSI H., ATIF S. Water Management, Impacts and Conflicts: Case of Indus water distribution in Sindh, Pakistan. International Journal of Rural studies, **19** (2), 3, **2012**.
25. GURU N., JHA R. Flood Frequency Analysis of Tel Basin of Mahanadi River System, India Using Annual Maximum and POT Flood Data. Aquatic Procedia, Elsevier, **4**, 427, **2015**.
26. 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**.
27. 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**.
28. ALAM M.A., EMURA K., FARNHAM C., YUAN J. Best-Fit Probability Distributions and Return Periods for Maximum Monthly Rainfall in Bangladesh. Climate, **6** (1), 1, **2018**.
29. CASSALHO F., BESKOW S., MELLO C.R. DE, MOURA M.M., KERSTNER L., ÁVILA L.F. At-Site Flood Frequency Analysis Coupled with Multiparameter Probability Distributions. Water Resources Management, **32** (1), 285, **2018**.
30. YUSOFF S.H.M., HAMZAH F.M., JAAFAR O. Multiparameter probability distributions of at-site Lmoment-based frequency analysis in Malaysia. International Journal of Mechanical Engineering, **7** (4), 48, **2022**.
31. HOSKING J.R.M. The theory of probability weighted moments. IBM Research Division, TJ Watson Research Center, pp. 1, **1986**.
32. 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: Statistical Methodology, **52** (1), 105, **1990**.
33. 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**.
34. HOSKING J.R.M., WALLIS J.R. Regional frequency analysis: an approach based on L-moments. Cambridge University Press, **2005**.
35. DAS S. Goodness-of-fit tests for generalized normal distribution for use in hydrological frequency analysis. Pure and Applied Geophysics, **175** (10), 3605, **2018**.
36. TEAM R.C. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria, **2022**.
37. ALI S., ARA J., SHAH I. A comparison of different parameter estimation methods for exponentially modified Gaussian distribution. Afrika Matematika, **33** (2), 1, **2022**.
38. 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**.
39. ZEINELDIN R.A., CHESNEAU C., JAMAL F., ELGARHI M. Statistical Properties and Different Methods of Estimation for Type I Half Logistic Inverted Kumaraswamy Distribution. Mathematics, **7** (10), 1, **2019**.
40. BERGMAN B. Estimation of Weibull parameters using a weight function. Journal of materials science letters, Springer, **5** (6), 611, **1986**.
41. ALQALLAF F., GHITANY M.E., AGOSTINELLI C. Weighted Exponential Distribution: Different Methods of Estimations. Applied Mathematics & Information Sciences, **1173** (3), 1167, **2015**.
42. DEY S., ALZAATREH A., GHOSH I. Parameter estimation methods for the Weibull-Pareto distribution. Computational and Mathematical Methods, **3** (1), e1053, **2021**.
43. 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**.
44. ZHAO X., ZHANG Z., CHENG W., ZHANG P.A. A New Parameter Estimator for the Generalized Pareto Distribution under the Peaks over Threshold Framework. Mathematics, **7** (5), 06, **2019**.
45. PABLO B.S., BRUCE E.R. Model Parameter Estimation Using Least Squares. Water Research, **26** (6), 789, **1992**.
46. 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**.
47. 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), 17, **2011**.
48. 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**.
49. MEYLAN P., FAVRE A.-C., MUSY A. Predictive hydrology: A frequency analysis approach, 2012, CRC Press **2012**.
50. KITE G.W. Frequency and Risk Analyses in Hydrology. Revised Ed, Water Resources Publications, LLC, **2019**.