Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Flood Frequency Analysis for Water Resources Planning and Management in Mahanadi River System, India

Rajesh Ranjan, Ramakar Jha*

Department of Civil Engineering, National Institute of Technology, Patna. India
Email: rj@nitp.ac.in

The research work aims to detailed flood frequency analysis (FFA) for water resources planning and management in the Mahanadi river system, India. Different commonly used flood frequency analysis techniques namely Normal, Gumbel max, Generalised Pareto (GP), Log-Pearson III (LP III), Log normal (LN), Log normal 3P(LN-3P), Log-Logistic(LL), Log-Logistic 3P(LL-3P), Pareto (P) and Generalised Extreme Value (GEV) method. For the study, the annual peak discharges of 19 gauging stations located in Mahanadi river system having length of data from 26 to 45 years have been used to estimate the flood frequency at return period of 5, 10, 25, 50, 100, 200, 500 and 1000 years. Amongst all the 19 gauging stations, the GEV distribution provides best results in many gauging stations followed by GP and LP-III distributions by considering three goodness of fit tests namely Kolmogorov-Smirov (KS), Anderson-Darling (AD), and Chi squared (CS) tests. At some gauging stations all the three goodness of fit test provides Rank 1, whereas at many stations KS and AD both goodness of fit tests provides Rank 1 in comparison to AD and CS tests. The annual maximum flow observed are compared well with the estimated annual maximum flow for the return period of 5, 10, and 25 years. Based on the results, it is observed that the GEV, GP and log-Pearson III distributions can be used for the design of water resources structure (culverts, canals, barrages, and dams) at different locations of the river system using 5, 10, 25, 50, 100, 200, 500 and 1000 years return period data.

1. Introduction

A Flood is an unusually high stage in a river, normally the level at which the river overflows its banks and inundates the adjoining area. Floods are an unfortunate occurrence, and consistent and exact stream flow forecasting is required for a variety of challenges such as water resource planning, strategy development, manoeuvring and maintenance activities.

Forecasting high-quality stream flow and making efficient use of this estimate provides significant financial and communal aid in water management. For the hydrologic constituent, interim and long-term stream flow forecasting is required for optimising systems or predicting future expansion or drop. Interim forecasting refers to hourly or day-to-day forecasting, which is critical for flood warning and safety, whereas long-term forecasting refers to monthly, seasonal, or annual timescales, which is very useful in reservoir processes and irrigation administration choices like distributing water to downstream consumers, arranging discharges, famine mitigation, and managing river agreements, or applying compacted acquiescence. The floods are resultants of a number of component parameters and are therefore very difficult to model analytically. This makes the estimation of the flood peak a very complex problem leading to many different approaches. One of the approach to the prediction of flood flows is the statistical method of frequency analysis.

Masmoudi and Habaieb (1993) developed statistical models, which were used on the Medjerdah River (Tunisia) to forecast dangerous flood occurrences. Model performance is described by statistical measures of accuracy, ultimate fault, and ultimate interruption among the measured and predicted flow with their alterations.

Griffis & Stedinger (2007) examined the characteristics of the Log-Pearson Type III (LP III) distribution in both real and logarithmic spaces. Their evaluation of U.S. flood data showed that the LP III distribution is a suitable model for annual flood records from natural, unregulated catchments, especially when considering skewness in logarithmic space. They also established relationships for the L-moment ratios of the LP III distribution, allowing for comparisons with regional statistical properties.

Rowinski et al. (2002) examined the Log-Gumbel and Log-Logistic distributions, two probability density functions frequently utilized in hydrological research. Their study addressed the application of these functions to hydrological data and the issues stemming from their mathematical characteristics. They highlighted that the maximum likelihood estimation method offers a way to align estimators beyond the valid range defined by the two L-moments

Rath et al. (2018) employed The Auto-Regressive Integrated Moving Average (ARIMA) model is utilized to predict the monthly mean inflow and daily inflow to the Hirakud Dam reservoir. Tools such as XLSTAT, STATA, and Microsoft Excel were employed for modeling ARIMA and validating the results. The approach involves short-term runoff forecasting on a yearly basis, where the predicted runoff data for one year is added to the observed dataset and used for forecasting runoff for the following year. Each forecasted value is subsequently treated as observed data for future predictions.

Helsel and Hirsch (1992) explored various probabilistic methods commonly employed in hydrology. They identified the Gumbel maximum value distribution and the Log-Pearson Type III (LP III) distribution as key models for analysing and solving water resource challenges.

Kamal et al. (2017) analysed discharge data from two locations and found that the Log-normal distribution is applicable for Haridwar, while the Gumbel EV1 distribution is suitable for Garhmukteshwar. Once an appropriate distribution is determined, it can be used to forecast discharge for specific return periods.

Brandimarte & Di Baldassarre (2012) introduced a method based on uncertain flood profiles to evaluate uncertainty in hydraulic modelling and frequency analysis (FFA). They systematically examined major sources of uncertainty to enhance the reliability of flood predictions.

Ewemoje & Ewemooje (2011) examines the use of Normal, Lognormal, and Log-Pearson Type 3 distributions for modelling at-site annual maximum flood flows, applying Hazen, Weibull, and California plotting positions in the Ogun-Oshun river basin, Nigeria. The Weibull plotting position, when paired with the Normal, Log-normal, and Log-Pearson Type III probability distributions, yielded the highest coefficient of determination values of 0.967, 0.987, and 0.986, respectively.

Mukherjee (2013) developed a model based on Gumbel's Extreme Value Distribution to predict the relationship between Peak Flood Discharge and Return Period. This model enables accurate estimation of Peak Flood Discharge for any specified return period (T) without requiring costly instrumentation or time-consuming fieldwork. The ability to estimate Peak Flood Discharge is essential for designing critical hydraulic structures such as Concrete Gravity Dams, Weirs, Barrages, Bridges, and Guide Banks. This model serves as a practical and efficient tool for ensuring the safety and functionality of such structures, particularly in flood-prone areas.

Subyani (2011) This study quantifies the hydrological characteristics and flood probabilities of key wadis in western Saudi Arabia, including Na'man, Fatimah, and Usfan. The flood responses of these wadis differ due to variations in their physical properties and rainfall distribution patterns.

Rainfall frequency analysis was carried out using the annual maximum 24-hour rainfall data from eight stations located across the region. The dataset spans 26 to 40 years, providing a comprehensive basis for analysis. Two widely applied statistical methods, Gumbel's Extreme Value Distribution and Log Pearson Type III Distribution, were used to evaluate the maximum daily rainfall and assess flood probabilities.

Sahoo et al. (2020) studies bivariate low-flow frequency analysis performed for the Mahanadi Basin, which exhibits notable variations in hydrological behaviour from upstream to downstream. The analysis focused on two key low-flow characteristics: severity (S) and occurrence (O). To estimate the joint return periods of low-flow events, three different copulas were tested for their applicability. The study analysed return periods of 2, 5, 10, 20, and 50 years, offering valuable insights into the frequency and intensity of low-flow events across the basin.

Pawar & Hire (2018) applied the Log Pearson Type III (LP-III) probability distribution to flood series data from four sites on the Mahi River—Mataji, Paderdi Badi, Wanakbori, and Khanpur—and three sites on its tributaries: Anas at Chakaliya, Som at Rangeli, and Jakham at Dhariawad. The annual maximum series data, covering a record length of 26 to 51 years, were analysed. Time series plots indicated that the Mahi River experienced its two largest recorded floods in 1973 and 2006.

Lima et al. (2016) study applies a multilevel hierarchical Bayesian framework to estimate local and regional Generalized Extreme Value (GEV) distribution parameters for flood frequency analysis, aiming to explicitly model and reduce uncertainties. In this framework, the GEV location and scale parameters for each site are assumed to follow independent log-normal distributions, with their mean parameters scaling with the drainage area. The shape parameter for each site is constrained (shrunk) towards a common mean based on empirical and theoretical considerations. Non-informative prior distributions are assigned to the hyperparameters, and the Markov Chain Monte Carlo (MCMC) method is employed to sample from the joint posterior distribution. The model is evaluated using annual maximum series data from 20 streamflow gauges across an 83,000 km² flood-prone basin in Southeast Brazil.

Bhat et al. (2019) study focuses on conducting a flood frequency analysis (FFA) of the River Jhelum in the Kashmir Basin. The Gumbel and Log-Pearson Type III (LP3) probability distributions were used to simulate future flood discharge scenarios based on annual peak flow data from three gauging stations on the River Jhelum (1956–2014). Design floods for various return periods (Tr) – 5, 10, 25, 50, 100, and 200 – were predicted and compared to evaluate potential future flood risks.

Guru & Jha (2014) studies at-site flood frequency analysis for estimating design floods, and selecting the appropriate probability distribution is essential for reliable flood frequency analysis. In this study, data from 19 flow gauging sites in the Mahanadi River Basin, India, were screened using the independence test. Seven probability distributions—Exponential, Gumbel, Logistic, Generalized Extreme Value (GEV), Generalized Pareto (GP), Pearson Type 3, and Lognormal—were evaluated to explore the suitability of various flood frequency models using the L-moment technique.

Tanaka et al. (2017) examined the impact of river overflow and dam operation of upstream areas on downstream extreme flood frequencies at Yodo River basin combining a flood-inundation model of upstream Kyoto City area with a rainfall-based flood frequency model and accounting for the probability of spatial and temporal rainfall pattern over the basin.

Here, various statistical methods are established for estimation of flow discharge at nineteen gauge stations in Mahanadi River basin, India. Also, goodness of fit is applied for analysing data sets.

2. THE STUDY AREA AND DATA COLLECTION

The Mahanadi River basin is the 8th largest basin and a major river of east central India, having total catchment area of 141,589 km^2 which is nearly 4.28% of the total geographical area of the country (Figure 1). The Mahanadi River basin extends over states of Chhattisgarh and Odisha and relatively smaller portions of Jharkhand, Maharashtra and Madhya Pradesh. The geographical extent of the basin lies between 80°28´ and 86°43´ east longitudes and 19°8´ and 23°32´ north latitudes. In the present study the river comprised of 310 km long from Hirakud Dam to the Naraj Gauging site, near Railway Bridge covering an area of 48,700 km^2.

Figure 1: The study area of Mahanadi river basin

The Mahanadi River begins at an elevation of approximately 442 meters above mean sea level, situated near the village of Pharsiya, in nearby to Nagri town located in the Raipur district of Chhattisgarh. The Mahanadi River spans a total length of about 851 km from its origin to its outfall into the Bay of Bengal, with 357 km flowing through Chhattisgarh and the remaining 494 km in Odisha. The table below shows the details of the catchment area, length and elevation at source of the important tributaries.(Table 1).

Table 1: The study area salient features

S.No.	Name of the Sub-basin	Bank	Elevation (msl)	Length (km)	Area (sq.km)	% Area
1	Mahanadi		442	851	48,230	34.1
2	Pairi	Right	488	113	3,503	2.5
3	Seonath	Left	533	383	30,761	21.7
4	Jonk	Right	762	196	3,673	2.6
5	Hasdeo	Left	915	333	9,803	6.9
6	Mand	Left	686	242	5,237	3.7
7	Ib	Left	762	251	12,447	8.8
8	Ong	Right	457	204	5,128	3.6
9	Tel	Right	700	296	22,818	16.1
			Total		1,41,600	100

The monsoon is the main rainy season for the Mahanadi basin, contributing more than 75% of the yearly rainfall. The Mahanadi basin receives an average annual rainfall of approximately 1400 mm. Due to its large geographic extent, the Mahanadi River basin exhibits significant spatial diversity in its hydro-meteorological features. The average daily temperature fluctuates between 13°C and 20°C in the winter and between 30°C and 37°C in the summer. The Mahanadi basin is delineated by the Central India Hills to the north, the Eastern Ghats to the south and east, and the Maikala Hill Range to the west. The Chiroli Hills act as the watershed, separating the Wainganga Valley from the Mahanadi basin, with its upper segment identified as the Chhattisgarh BasinFigure 2 illustrates the Annual Maximum discharge of all 19 stations considered in the present work.

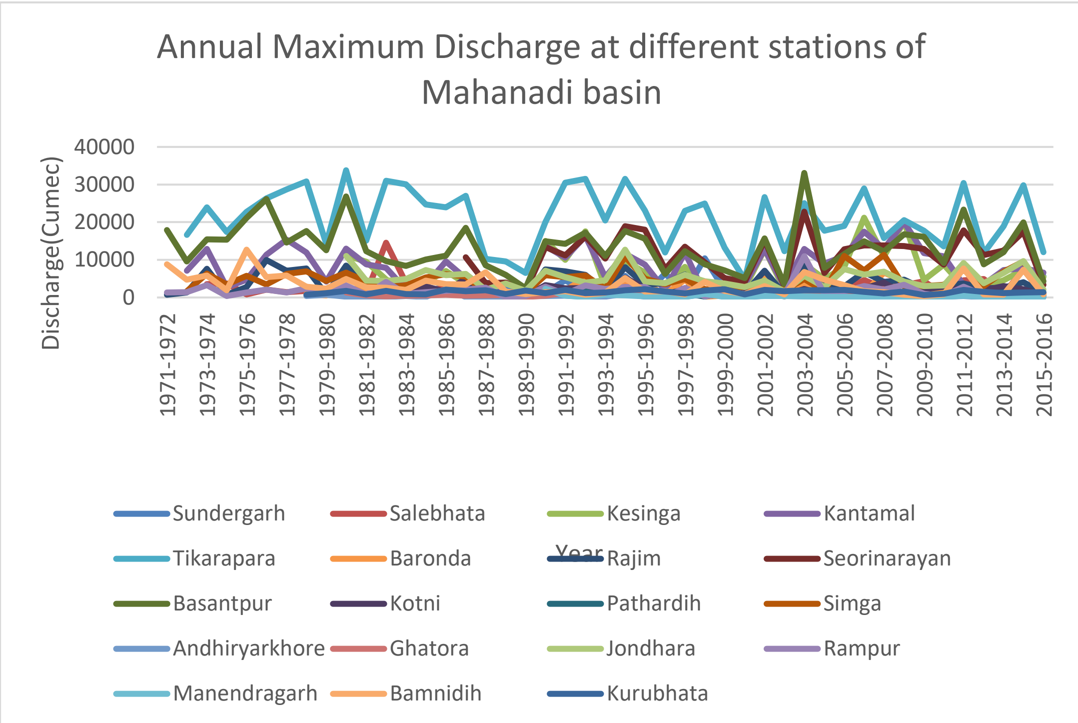


Figure 2: Annual maximum discharge of all the stations lying in the study area

3. METHODOLOGY

The value of the annual maximum flood from a given catchment area for large number of successive years constitute a hydrologic data series called the annual series. In this study an exhaustive & detailed Flood frequency analysis is done using ten statistical distribution i.e Normal, Gumbel max, Generalised Pareto (GP), Log-Pearson III (LP III), Log normal (LN), Log normal 3P(LN-3P), Log-Logistic(LL), Log-Logistic 3P(LL-3P), Pareto (P) and Generalised Extreme Value (GEV) method for all the nineteen gauging station to determine the maximum flood discharge for different return period using Annual Maximum Discharge Data annual series. Each method predicts the flood peak with certain advantages and disadvantages. These methods require mean, standard deviation, skewness coefficient, kurtosis coefficient and return period value which is computed using the annual flood series data. In frequency analysis of floods the usual problem is to predict extreme flood events. Towards this, specific extra-value distributions are assumed and the required statistical parameters calculated from the available data. Using this the flood magnitude for a specific return period is estimated. Chow (1951) has shown that most frequency distribution functions applicable in hydrologic studies can be expressed by the following equation known as the general equation of hydrologic frequency analysis:

$$Q_p = \mu + K_t \sigma \quad (1)$$

Where Q_p= value of the variate Q of a random hydrologic series with a return period T, μ = mean of the variate, σ=standard deviation of the variate, K_t= Frequency factor which depends upon the return period, T and the assumed frequency distribution. Table 2 shows governing equations used in different statistical methods.

Table 2: Governing equations used in different methods

Distributions	Probability Density Function	Cumulative Distribution Function
Normal	$f(x) = \frac{\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)}{\sigma\sqrt{2\pi}}$ (2)	$F(x) = \varphi\left(\frac{x-\mu}{\sigma}\right)$ (3)
Log Normal	$f(x) = \frac{\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)}{x\sigma\sqrt{2\pi}}$ (4)	$F(x) = \varphi\left(\frac{\ln x-\mu}{\sigma}\right)$ (5)
Gumbel Max	$f(x) = \frac{1}{\sigma}\exp(-z-\exp(-z))$ (6) $z = \frac{x-\mu}{\sigma}$	$F(x) = \exp(-\exp(-z))$ (7)
Generalized Extreme Value	$f(x) = \begin{cases} \frac{1}{\sigma}\exp\left(-(1+kz)^{\frac{-1}{k}}\right)(1+kz)^{-1-\frac{1}{k}} & ; k \neq 0 \\ \frac{1}{\sigma}\exp(-z-\exp(-z)) & ; k = 0 \end{cases}$ (8)	$F(x) = \begin{cases} \exp\left(-(1+kz)^{\frac{-1}{k}}\right) & ; k \neq 0 \\ \exp(-\exp(-z)) & ; k = 0 \end{cases}$ (9)
Log Logistic	$f(x) = \frac{(\beta/\alpha)(x/\alpha)^{\beta-1}}{(1+(x/\alpha)^{\beta})^2}, x > 0$ (10)	$F(x) = \frac{1}{1+(x/\alpha)^{-\beta}}, x > 0$ (11)
Generalised Pareto	$f(x) = \begin{cases} \frac{1}{\sigma}\left(1 + k\,\frac{x-\mu}{\sigma}\right)^{-\frac{1}{k}-1} & ; k \neq 0 \\ \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\right) & ; k = 0 \end{cases}$ (12)	$F(x) = \begin{cases} 1-(1+k\frac{x-\mu}{\sigma})^{-\frac{1}{k}} & ; k \neq 0 \\ 1-\exp\left(-\frac{x-\mu}{\sigma}\right) & ; k = 0 \end{cases}$ (13)
Log Pearson Type III	$f(x) = \frac{1}{x\|\beta\|\Gamma(\alpha)}\left(\frac{\ln x-\gamma}{\beta}\right)^{\alpha-1}\exp\left(-\frac{\ln x-\gamma}{\beta}\right)$ (14)	$F(x) = \frac{\Gamma_{\frac{(\ln x-\gamma)}{\beta}}(\alpha)}{\Gamma(\alpha)}$ (15)
Log Logistic (3P)	$f(x) = \frac{(x-\gamma/\alpha)^{\alpha-1}}{\beta[\left(1+\left(x-\frac{\gamma}{\beta}\right)^{\alpha}\right)]^2},$ (16)	$F(x) = \frac{1}{1+(\frac{\beta}{x-\gamma})^{\alpha}}$ (17)
Log Normal (3P)	$f(x) = \frac{\exp\left(-\left(\frac{\ln(x-\gamma)-\mu)}{2\sigma 2}\right)^2\right)}{(x-\gamma)\sigma\sqrt{2\pi}}$ (18)	$F(x) = \varphi\left(\frac{\ln(x-\gamma)-\mu}{\sigma}\right)$ (19)
Pareto	$F(x) = \frac{\alpha x_m^{\alpha}}{x^{\alpha+1}}$ (20)	$F(x) = 1-(\frac{x_m}{x})^{\alpha}$ (21)
(k, α) are shape, (σ,β) are scale and (μ,γ) are location parameter;φ=Laplace Integral; Γ=gamma function		

3.1 Goodness of fit tests

For a given set of data, whether a certain distribution is fit or not is checked using this test. Quality of fit for the observed data set is ranked through calculation of statistical parameters. Affinity of samples from the expected theoretical probability distribution is assessed. To evaluate null hypothesis, it is applied and discarded if the observed test surpasses the critical value for the constant significance level. Chi-squared, Anderson–Darling (AD) and Kolmogorov–Smirnov (KS) tests are employed here.

3.1.1 Kolmogorov–Smirnov test

Discovering whether a sample is from an assumed continuous probability distribution is the main objective of this test. It is on the basis of empirical cumulative distribution functions (CDF), that is:

$$F_m(y) = \frac{1}{m} \times [\text{Observation number } \leq y] \tag{22}$$

The Kolmogorov–Smirnov test statistic (K) is given by prevalent perpendicular difference in hypothetical and experiential CDF:

$$K = \max_{1 \leq j \leq m} \left(F(y_i) - \frac{j-1}{m}, \frac{j}{m} - F(y_j)\right) \tag{23}$$

3.1.2 Anderson–Darling test

This associates the fit of an observed to an expected CDF, hence giving additional weight to distribution tails compared to previous experiments.

$$D^2 = -m - \frac{1}{m}\sum_{j=1}^{m}(2j-1) \times \left[\ln F(y_j) + \ln\left(1 - F(y_{m-j+1})\right)\right] \tag{24}$$

3.1.3 Chi-squared test

This is applied to find out whether a sample has come from a population with a given distribution. Binned data are applied, and hence the value of the test statistic depends on how data are binned.

$$X^2 = \sum_{j=1}^{l} \frac{(O_j - E_j)^2}{E_j} \tag{25}$$

Where, O_j = observed frequency; j = observations' number; Expected frequency (E_j) = F(Y2) - F(Y1); F = ???????????????????????????????????????

The cumulative distribution function is

$$l = 1 + \log_2 m \tag{27}$$

Where m = sample size.

4. RESULTS AND DISCUSSIONS

The described methodology is utilized to calculate statistical parameters, which are then used with frequency distribution techniques to determine the design discharge for return periods of

5, 10, 25, 50, 100, 200, 500, and 1000 years. The shape, scale, and location parameters for each distribution are presented in Table 3.

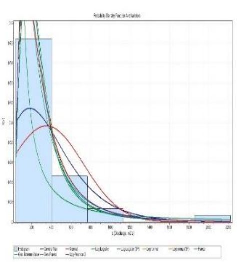
Table 3: Parameters estimated in different methods used in the present owrk

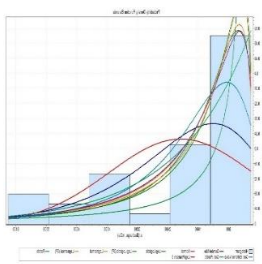
FFA	Normal	Gumbel Max	Gen. Pareto	Log-Pearson 3	Lognormal	Lognormal (3P)	Gen. Extreme Value
Andhiyarkhore	σ=374.97 μ=349.51	σ=292.37 μ=180.75	k=0.23685 σ=209.15 μ=75.447	α=2439.7 β=0.01646 γ=-34.633	σ=0.80216 μ=5.5206	σ=0.7885 μ=5.5376 γ=-3.1242	k=0.3914 σ=132.81 μ=190.1
Bamnidhi	σ=2662.2 μ=3229.1	σ=2075.7 μ=2030.9	k=-0.06713 σ=3119.2 μ=306.06	α=1022.2 β=-0.02672 γ=35.06	σ=0.84476 μ=7.7462	σ=0.92272 μ=7.6543 γ=138.35	k=0.19867 σ=1636.5 μ=1888.6
Baronda	σ=1886.7 μ=2048.2	σ=1471.0 μ=1199.0	k=0.02432 σ=1929.7 μ=70.401	α=432.72 β=-0.04769 γ=27.824	σ=0.97888 μ=7.1879	σ=1.0976 μ=7.0685 γ=87.412	k=0.25438 σ=1072.5 μ=1073.0
Basantpur	σ=6380.8 μ=13420.0	σ=4975.1 μ=10549.0	k=-0.564 σ=14199.0 μ=4341.3	α=6.1999 β=-0.21496 γ=10.714	σ=0.52927 μ=9.381	σ=0.26807 μ=10.016 γ=-9775.5	k=-0.07546 σ=5452.0 μ=10654.0
Ghatora	σ=496.91 μ=656.82	σ=387.44 μ=433.18	k=0.01358 σ=454.23 μ=196.34	α=161.48 β=-0.05223 γ=14.709	σ=0.65438 μ=6.2755	σ=0.58928 μ=6.3771 γ=-46.689	k=0.24774 σ=250.77 μ=431.68
Jondhara	σ=496.91 μ=656.82	σ=387.44 μ=433.18	k=0.01358 σ=454.23 μ=196.34	α=161.48 β=-0.05223 γ=14.709	σ=0.65438 μ=6.2755	σ=0.58928 μ=6.3771 γ=-46.689	k=0.24774 σ=250.77 μ=431.68
Kantamal	σ=4938.0 μ=8259.6	σ=3850.1 μ=6037.2	k=-0.77371 σ=13970.0 μ=383.4	α=7.2354 β=-0.2888 γ=10.869	σ=0.76796 μ=8.7793	σ=0.30036 μ=9.6776 γ=-8417.2	k=-0.17892 σ=4714.0 μ=6258.3
Kesinga	σ=5327.0 μ=6930.6	σ=4153.4 μ=4533.1	k=-0.17674 σ=7447.6 μ=601.5	α=21.989 β=-0.18532 γ=12.6	σ=0.85685 μ=8.5247	σ=0.75112 μ=8.6514 γ=-487.36	k=0.13424 σ=3645.8 μ=4272.5
Kotni	σ=1200.0 μ=2002.1	σ=935.62 μ=1462.0	k=-0.43937 σ=2317.4 μ=392.11	α=12.87 β=-0.18763 γ=9.8229	σ=0.66314 μ=7.4081	σ=0.44904 μ=7.7795 γ=-636.22	k=-0.0108 σ=961.63 μ=1457.2
Kurubhata	σ=466.79 μ=1477.6	σ=363.96 μ=1267.5	k=-1.2478 σ=1981.2 μ=596.18	α=12.924 β=-0.09729 γ=8.5001	σ=0.34514 μ=7.2427	σ=0.04468 μ=9.2472 γ=-8909.7	k=-0.39146 σ=503.56 μ=1333.2
Manendragarh	σ=450.94 μ=368.71	σ=351.59 μ=165.77	k=0.44678 σ=153.77 μ=90.758	α=4.2735 β=0.37477 γ=3.9478	σ=0.75969 μ=5.5494	σ=1.1949 μ=4.9867 γ=75.897	k=0.53852 σ=110.98 μ=179.31
Pathardih	σ=439.18 μ=1068.1	σ=342.43 μ=870.46	k=-1.1175 σ=1672.3 μ=278.36	α=2.6042 β=-0.32626 γ=7.7137	σ=0.51628 μ=6.8641	σ=0.04407 μ=9.1862 γ=-8703.4	k=-0.33572 σ=458.88 μ=921.45
Rajim	σ=2781.5 μ=3673.1	σ=2168.8 μ=2421.2	k=-0.3953 σ=5271.6 μ=-105.02	α=11.321 β=-0.29385 γ=11.154	σ=0.97764 μ=7.8274	σ=0.79389 μ=8.0263 γ=-367.26	k=0.01268 σ=2248.7 μ=2346.5
Rampur	σ=1749.8 μ=1835.2	σ=1364.3 μ=1047.7	k=-0.11511 σ=1884.6 μ=145.16	α=7.8422 β=-0.34716 γ=9.8537	σ=0.96133 μ=7.1312	σ=0.71416 μ=7.4003 γ=-267.67	k=0.17016 σ=959.19 μ=1089.3
Salebhata	σ=2551.9 μ=2688.9	σ=1989.7 μ=1540.4	k=0.04237 σ=2249.1	α=13.328 β=-0.25271	σ=0.9118 μ=7.5295	σ=0.75698 μ=7.697 γ=-	k=0.2656 σ=1264.4 μ=1514.2

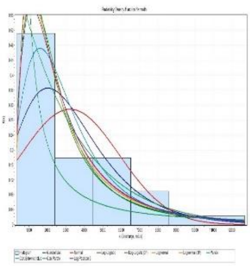
			μ=340.27	γ=10.898		234.93	
Seorinarayan	σ=5448.1 μ=11118.0	σ=4247.9 μ=8665.8	k=-1.0977 σ=20437.0 μ=1374.9	α=3.5066 β=-0.3486 γ=10.374	σ=0.64181 μ=9.1513	σ=0.04883 μ=11.611 γ=-99253.0	k=-0.32706 σ=5674.3 μ=9274.2
Simga	σ=2666.2 μ=4669.8	σ=2078.8 μ=3469.9	k=-0.44238 σ=5238.1 μ=1038.2	α=16.775 β=-0.15489 γ=10.871	σ=0.62714 μ=8.2726	σ=0.44198 μ=8.6085 γ=-1355.7	k=-0.01239 σ=2169.6 μ=3443.8
Sundergarh	σ=1950.7 μ=2336.2	σ=1521.0 μ=1458.3	k=0.34497 σ=961.57 μ=868.22	α=3.6244 β=0.32099 γ=6.3753	σ=0.60301 μ=7.5387	σ=1.0471 μ=6.8398 γ=731.12	k=0.46559 σ=652.64 μ=1409.2
Tikarapara	σ=7688.4 μ=21101.0	σ=5994.6 μ=17641.0	k=-1.2447 σ=32479.0 μ=6631.8	α=3.1727 β=-0.24934 γ=10.665	σ=0.43905 μ=9.8744	σ=0.03506 μ=12.292 γ=-1.9696E+5	k=-0.39017 σ=8270.0 μ=18724.0

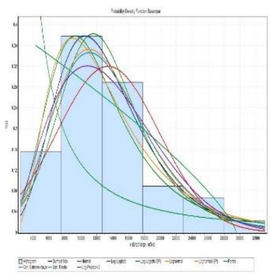
The functions listed above, combined with the fitting parameter values summarized in Table 3, are used to plot the PDF and CDF graphs for each frequency distribution method.

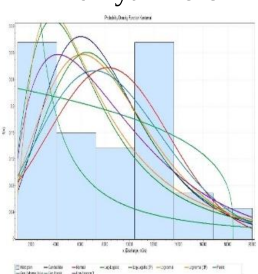
The Probability density function is a function that provides the likelihood that the value of a random variable will fall between a certain range of values. The graph of a probability density function is in the form of a bell curve. The area that lies between any two specified values gives the probability of the outcome of the designated observation. The term cumulative distribution function or CDF is a method to describe the distribution of random variables. This random variable may be discrete, continuous, or mixed. It is the probability function that gives the probability that a random variable x is less than or equal to the independent variable of the function. Figure 3 and Figure 4 shows the PDF and CDF graphs at all 19 stations of Mahanadi river system respectively.

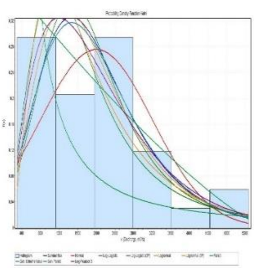

Andhiyarkhore

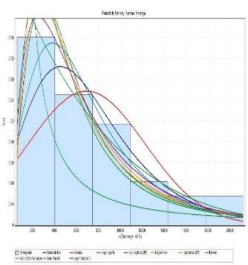

Baronda

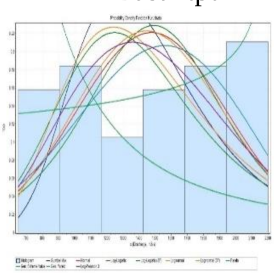

Bamnidih

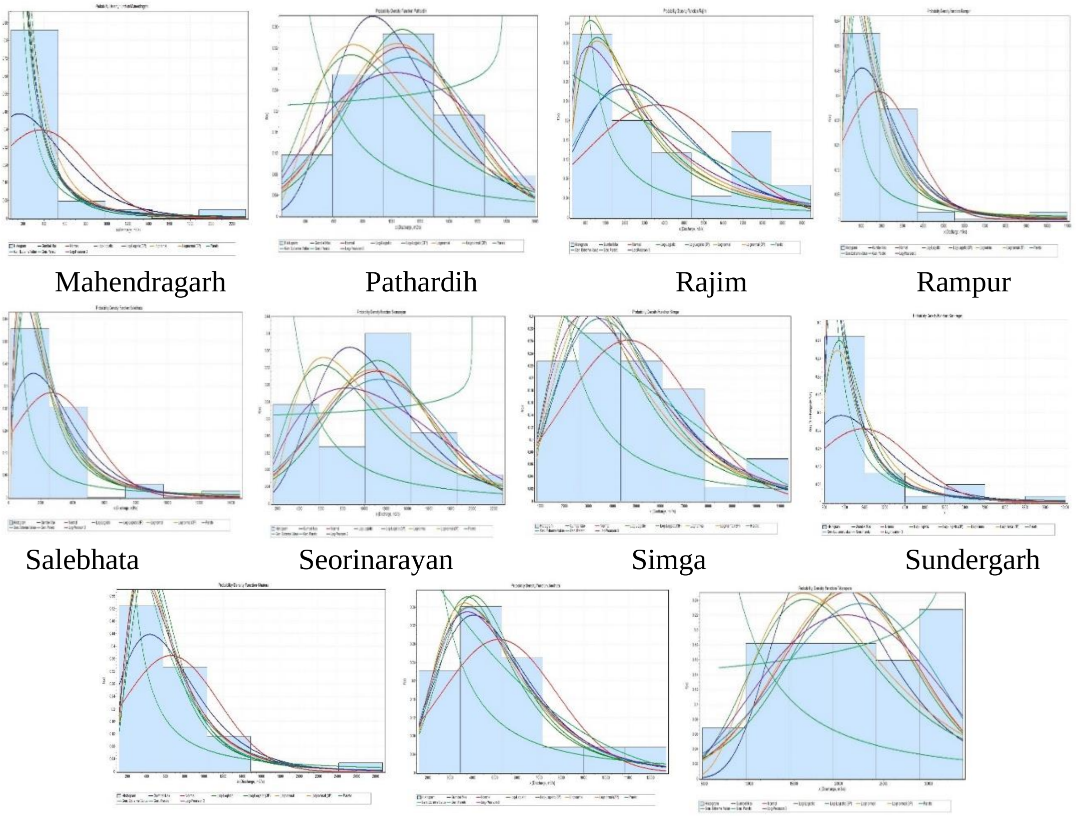

Basantpur


Kantamal


Kotni


Kesinga


Kurubhata

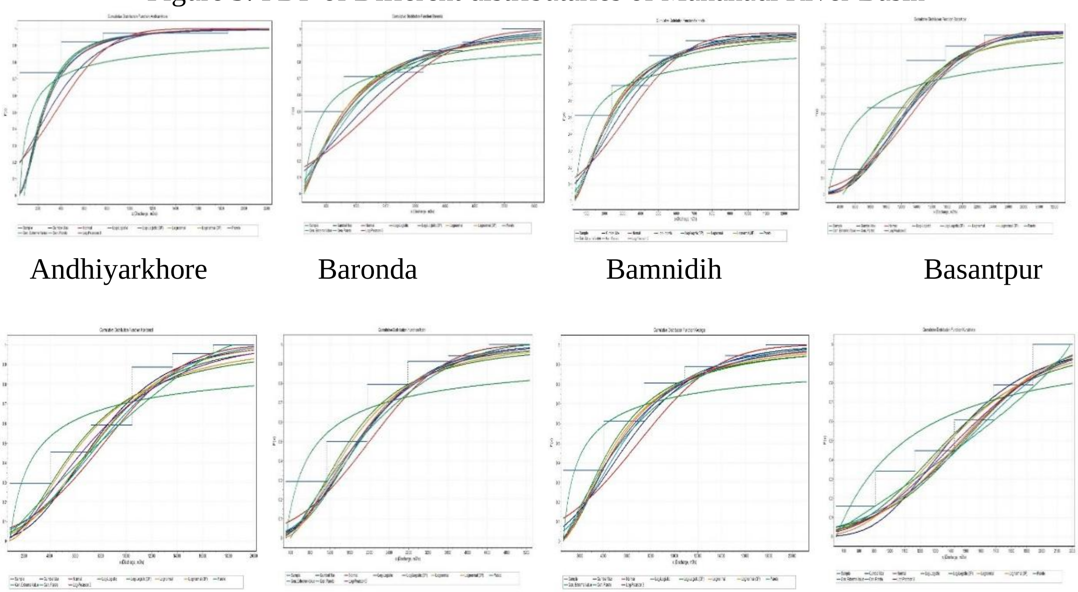


Mahendragarh Pathardih Rajim Rampur

Salebhata Seorinarayan Simga Sundergarh

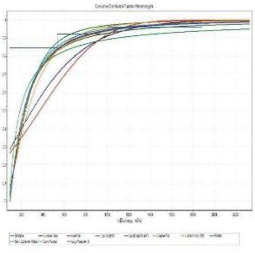
Ghatora Jondhara Tikarapara

Figure 3: PDF of Different distributaries of Mahanadi River Basin

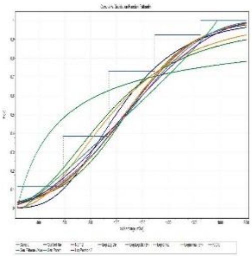


Andhiyarkhore Baronda Bamnidih Basantpur

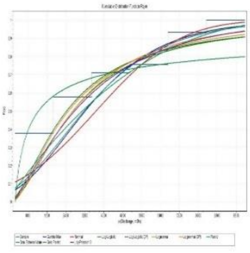
Kantamal Kotni Kesinga Kurubhata



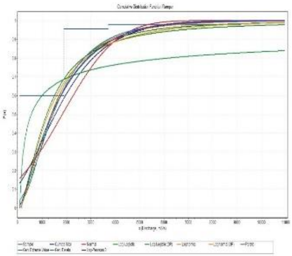
Mahendragarh



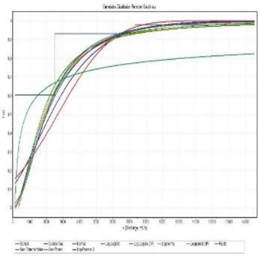
Pathardih



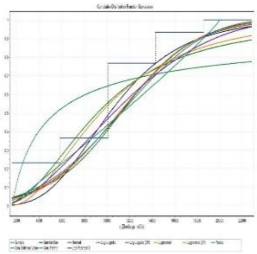
Rajim



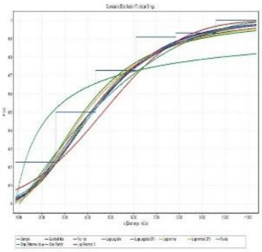
Rampur



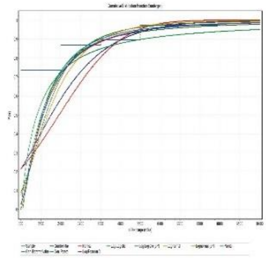
Salebhata



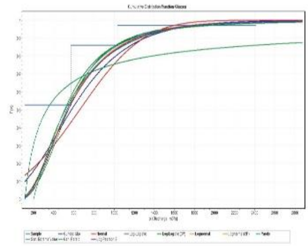
Seorinarayan



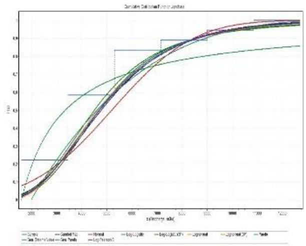
Simga



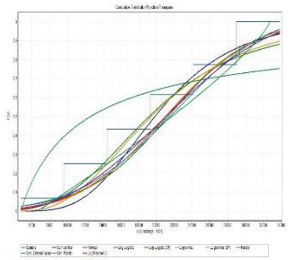
Sundergarh



Ghatora



Jondhara



Tikarapara

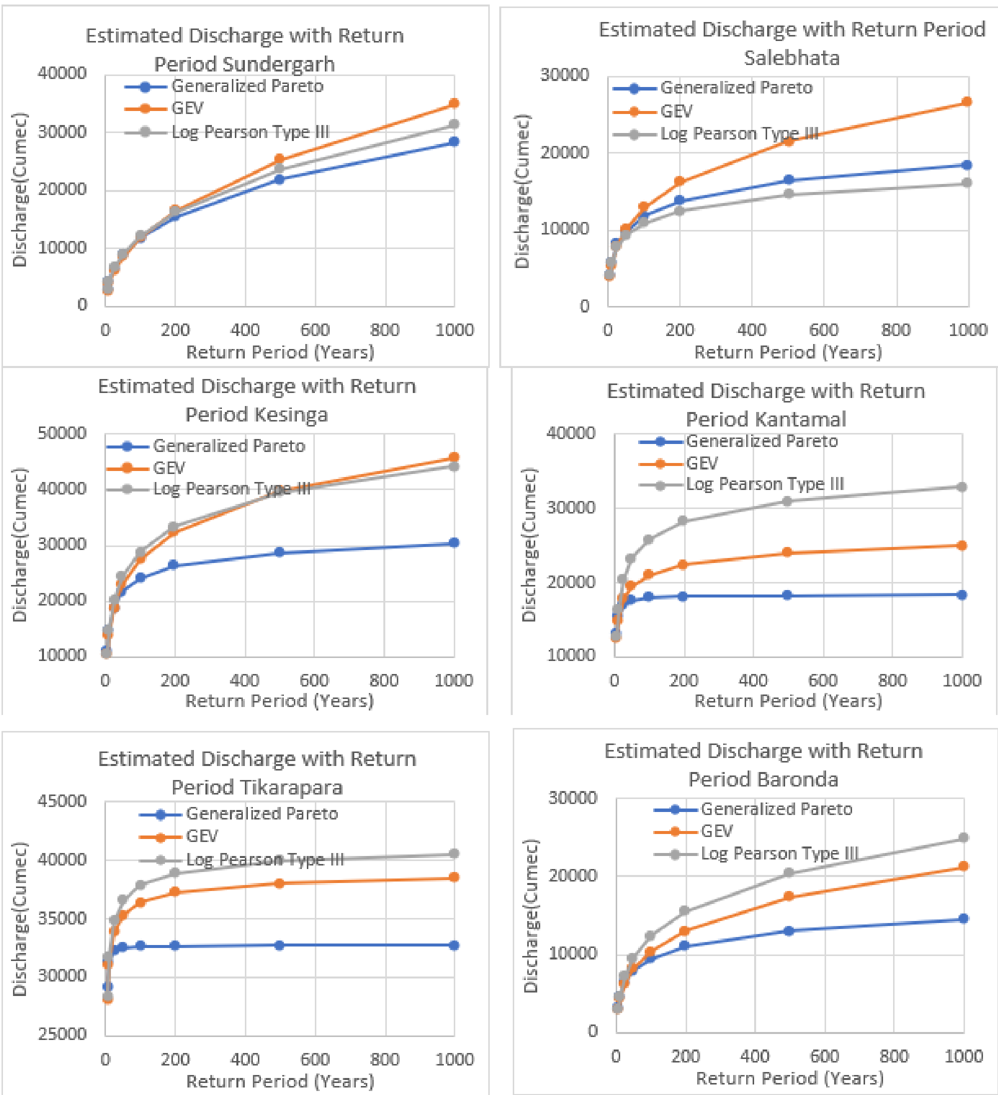
Figure 4: CDF of Different distributaries of Mahanadi River Basin

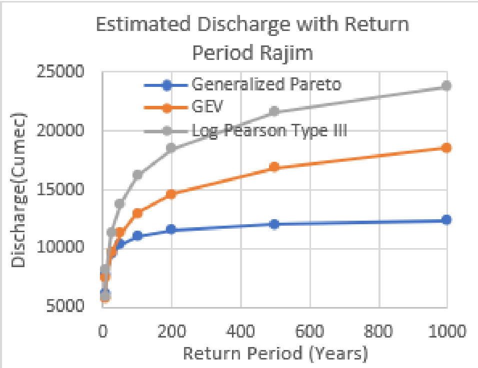
The Goodness of fit shows that as per Kolmogorov–Smirov Goodness of fit test GEV is best for Andhiyarkhore, Basantpur, Ghatora, Jondhara Kotni, Pathardih , Rampur & Seorinarayan, Gen. Pareto is best for Kantamal, Kesinga, Kurubhata, Tikarapara, Rajim, Baronda & Sundergarh, Log Pearson 3 is best for Salebhata, Simga, Manendragarh & Bamnidih. As per Anderson–Darling Goodness of fit test GEV is best for Andhiyarkhore, Basantpur, Ghatora, Jondhara, Kotni, Pathardih , Rampur, Kantamal, Kurubhata, Salebhata, Sundergarh & Seorinarayan, Gen. Pareto is best for Bamnidih & Rajim, Log-Pearson 3 is best for Baronda, Kesinga, Manendragarh, Simga & Tikarapara, As per Chi-Suuared Goodness of fit test GEV is best for Kurubhata, Pathardih, Rampur & Sundergarh, Gen. Pareto is best for Bamnidih & Rajim, Gumbel Max is best for Simga, Log Pearson 3 is best for Basantpur, Ghatora, Jondhara, Kantamal, Kotni, Salebhata & Tikarapara, Lognormal is best for Andhiyakhore, Baronda, Kesinga, Manendragarh & Seorinarayan.

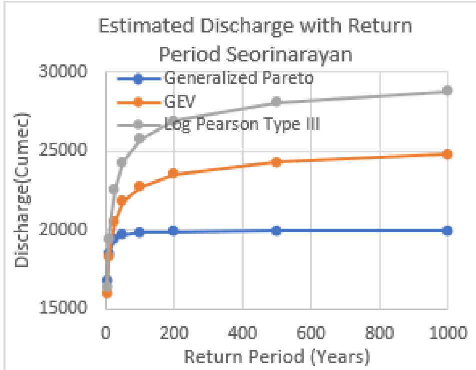
It is observed that for the low discharge value LPIII is providing the best result(Below the Hirakud Dam), Kesinga(Tel), Kantamal(Tel), Salebhata(Ong) & Tikarapara showing good result in LP3 as per goodness of fit test. Based on Goodness of fit test GEV is showing best result in upstream of basin at 12 station i.e on Andhiyarkore, Kotni, Pathardhi, Jondhara, Rampur, Seorinarayan, Ghatora, Basantpur, Kurubhata, Sundergarh, Salebhata, Kantamal based on the as per either 1, 2 or 3 (KS, AD & CS) goodness of fit test. In GP best result has been shown for Baronda & Rajim station due to regulated flow made up with structured weir/dam/barrage in upstream of it.

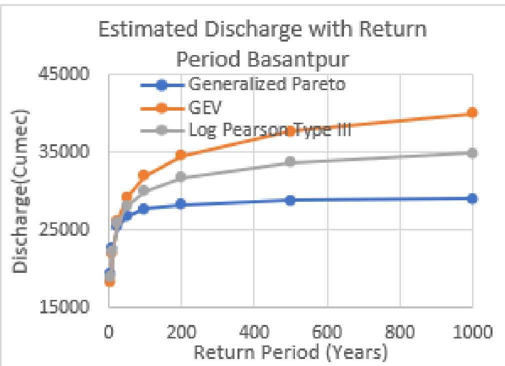
4.1 Annual Maximum Flood (AMF) at different return period

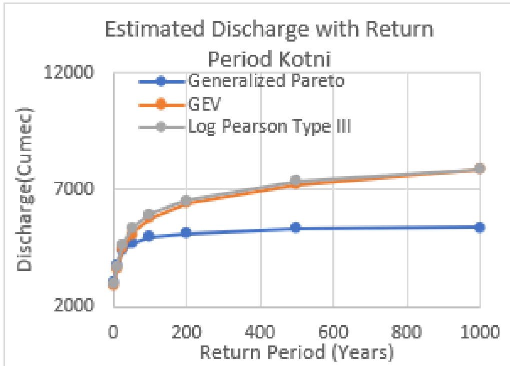
From the results, the Generalised Extreme Value (GEV) is found to be best among all other flood frequency distribution models. The Generalised Pareto (GP) and Long-Pearson Type-III are found to be at Ranl 2 and Rank 3 for flood frequency analysis. Keeping this in view, prediction of floods at all sampling stations of river Mahanadi has been done using all the three methods for different return periods. Figure 5 illustrates the values of Annual Maximum Flood at different return periods at all the stations.

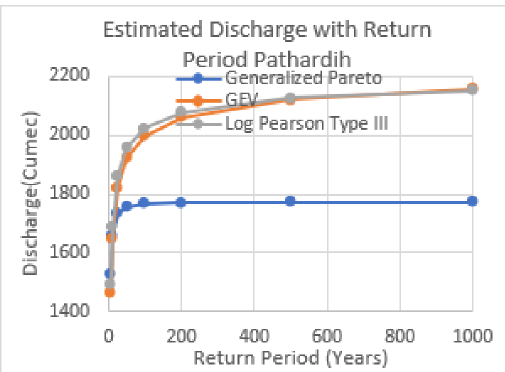


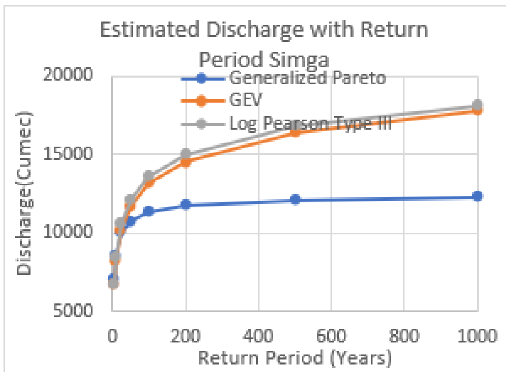

Estimated Discharge with Return Period Rajim
25000
20000
15000
10000
5000
Discharge(Cumec)
Generalized Pareto
GEV
Log Pearson Type III
0
200
400
600
800
1000
Return Period (Years)

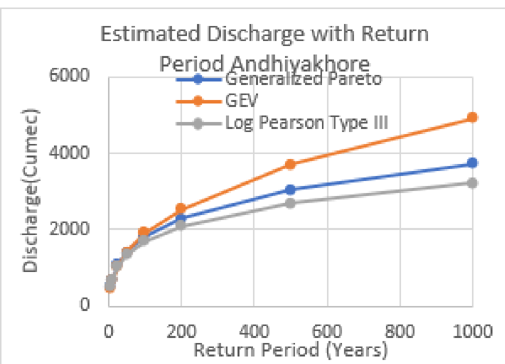

Estimated Discharge with Return Period Seorinarayan
30000
25000
20000
15000
Discharge(Cumec)
Generalized Pareto
GEV
Log Pearson Type III
0
200
400
600
800
1000
Return Period (Years)

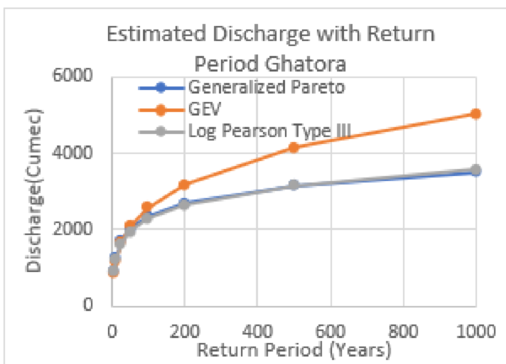

Estimated Discharge with Return Period Basantpur
45000
35000
25000
15000
Discharge(Cumec)
Generalized Pareto
GEV
Log Pearson Type III
0
200
400
600
800
1000
Return Period (Years)


Estimated Discharge with Return Period Kotni
12000
7000
2000
Discharge(Cumec)
Generalized Pareto
GEV
Log Pearson Type III
0
200
400
600
800
1000
Return Period (Years)


Estimated Discharge with Return Period Pathardih
2200
2000
1800
1600
1400
Discharge(Cumec)
Generalized Pareto
GEV
Log Pearson Type III
0
200
400
600
800
1000
Return Period (Years)


Estimated Discharge with Return Period Simga
20000
15000
10000
5000
Discharge(Cumec)
Generalized Pareto
GEV
Log Pearson Type III
0
200
400
600
800
1000
Return Period (Years)


Estimated Discharge with Return Period Andhiyakhore
6000
4000
2000
0
Discharge(Cumec)
Generalized Pareto
GEV
Log Pearson Type III
0
200
400
600
800
1000
Return Period (Years)


Estimated Discharge with Return Period Ghatora
6000
4000
2000
0
Discharge(Cumec)
Generalized Pareto
GEV
Log Pearson Type III
0
200
400
600
800
1000
Return Period (Years)

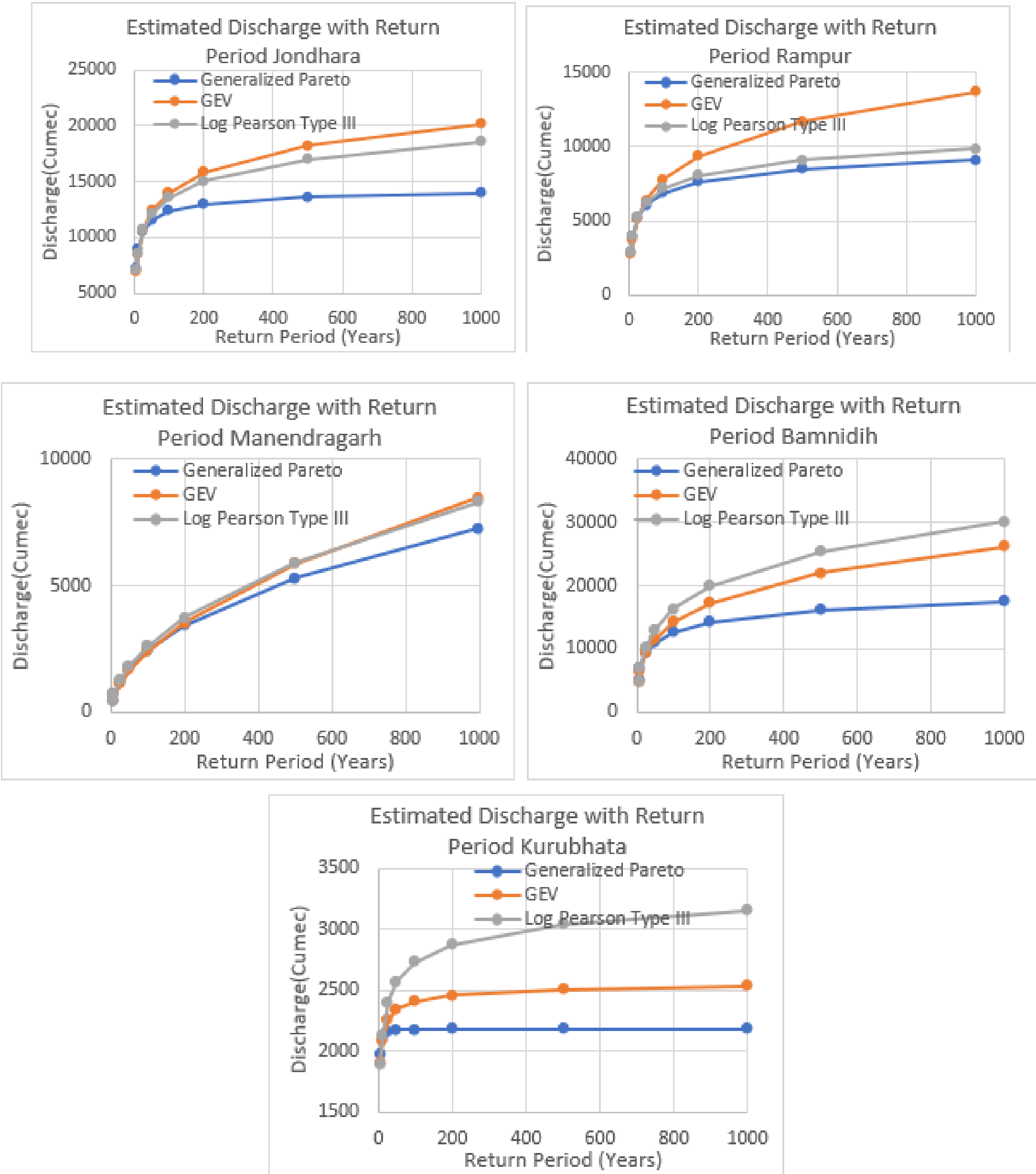


Figure 5: Annual Maximum Flood prediction using GEV, GP and LP-III

From Figure 5, it is observed that the annual maximum discharge at 5, 10, -------- year return period are observed to be highest at --------- station and lowest at ----------- station respectively. The estimated values are useful for any kind of water storage structures at different location.

5. Conclusions

In this paper, an effort has been made to forecast discharges at various return periods using statistical methods. A total of Ten statistical methods have been used to predict annual maximum discharge in the Mahanadi River basin, covering nineteen stations. The rate of increase of annual maximum discharge is very high at the initial return periods and then the rate of increase eventually lower. In most of the cases, GEV gives the ~~peak flood discharge~~ annual maximum flood and GP distribution contributes to the least discharge. For design of any water resources structures, the results obtained using GEV should be preferred for safe structural design. The annual maximum flood values obtained for 5, 10, 20, -------- year return period can used for design of culverts, minor bridges, canal, major bridges, weir, barrages, and Dams at different locations in coming years for water storage and management.

The influencing factor of frequency is analysed on the basis of runoff complexity from drainage basins. It is found that flow probability increases at the upstream of Mahanadi, which may be characterized by the underlying surface condition change influenced by human activities and geomorphology changes, and be considered for future scope. In other sections, the purpose of the research is to diminish future flood damage in the river basin. Hence, forecast of flow discharge is a key indication towards hydrological modelling and development for water resources engineering.

References

1. Masmoudi, M. & Habaieb, H. 1993 The performance of some real-time statistical flood forecasting models seen through multicriteria analysis. Water Resources Management 7 (1), 57–67
2. Griffis, V. W. & Stedinger, J. R. 2007 Log-Pearson type 3 distribution and its application in flood frequency analysis. II: parameter estimation methods. Journal of Hydrologic Engineering 12 (5), 492–500
3. Rowinski, P. M., Strupczewski, W. G. & Singh, V. P. 2002 A note on the applicability of log-Gumbel and log-logistic probability distributions in hydrological analyses: I. Hydrological Sciences Journal 47 (1), 107–122
4. Rath, A., Samantaray, S., Bhoi, K. S. & Swain, P. C. 2018 Flow forecasting of hirakud reservoir with ARIMA model. In: 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, India, pp. 2952–2960.
5. Helsel, D. R. & Hirsch, R. M. 1992 Statistical Methods in Water Resources, 1st edn, Studies in Environmental Science, Vol. 49. Elsevier
6. Kamal, V., Mukherjee, S., Singh, P., Sen, R., Vishwakarma, C. A., Sajadi, P., Asthana, H. & Rena, V. 2017 Flood frequency analysis of Ganga river at Haridwar and Garhmukteshwar. Applied Water Science 7 (4), 1979–1986.
7. Brandimarte, L. & Di Baldassarre, G. 2012 Uncertainty in design flood profiles derived by hydraulic modelling. Hydrology Research 43 (6), 753–761.
8. Ewemoje, T. A. & Ewemooje, O. S. 2011 Best distribution and plotting positions of daily maximum flood estimation at Ona River in Ogun-Oshun river basin. Nigeria. Agricultural Engineering International: CIGR Journal 13 (3), 1–11.
9. Mukherjee, M. K. 2013 Flood frequency analysis of River Subernarekha, India, using Gumbel's extreme value distribution. International Journal of Computational Engineering Research 3 (7), 12–19.
10. Subyani, A. M. 2011 Hydrologic behavior and flood probability for selected arid basins in

Makkah area, western Saudi Arabia. Arabian Journal of Geosciences 4 (5–6), 817–824.

11. Sahoo, B. B., Jha, R., Singh, A. & Kumar, D. 2020 Bivariate low flow return period analysis in the Mahanadi River basin, India using copula. International Journal of River Basin Management 18, 107–116.
12. Pawar, U. & Hire, P. 2018 Flood frequency analysis of the Mahi Basin by using Log Pearson Type III probability distribution. Hydrospatial Analysis 2 (2), 102–112.
13. Lima, C. H., Lall, U., Troy, T. & Devineni, N. 2016 A hierarchical Bayesian GEV model for improving local and regional flood quantile estimates. Journal of Hydrology 541, 816–823.
14. Bhat, M. S., Alam, A., Ahmad, B., Kotlia, B. S., Farooq, H., Taloor, A. K. & Ahmad, S. 2019 Flood frequency analysis of river Jhelum in Kashmir basin. Quaternary International 507, 288–294.
15. N. Guru & R. Jha,2014, A study on selection of probability distributions for at-site flood frequency analysis in Mahanadi River Basin, India.
16. Tanaka, T., Tachikawa, Y., Ichikawa, Y. & Yorozu, K. 2017 Impact assessment of upstream flooding on extreme flood frequency analysis by incorporating a flood-inundation model for flood risk assessment. Journal of Hydrology 554, 370–382.