Estimation of flood frequency using statistical method: Mahanadi river basin, India

Sandeep Samantaray* and Abinash Sahoo

Department of Civil Engineering, NIT Silchar, Assam, India

*Corresponding author. E-mail: samantaraysandeep963@yahoo.com

Abstract

Estimating stream flow has a substantial financial influence, because this can be of assistance in water resources management and provides safety from scarcity of water and conceivable flood destruction. Four common statistical methods, namely, Normal, Gumbel max, Log-Pearson III (LP III), and Gen. extreme value method are employed for 10, 20, 30, 35, 40, 50, 60, 70, 75, 100, 150 years to forecast stream flow. Monthly flow data from four stations on Mahanadi River, in Eastern Central India, namely, Rampur, Sundargarh, Jondhra, Basantpur are used in the study. Result shows that Gumbel max gives better value of flow discharge than the Normal, LP III, and Gen. extreme value methods for all four gauge stations. Estimated flood values for Rampur, Sundargarh, Jondhra, Basantpur stations are 372.361 m³/sec, 530.415 m³/sec, 2,133.888 m³/sec, 3,836.22 m³/sec, respectively, considering Gumbel max. Goodness of fit test for four statistical distribution techniques applied in the present study are also evaluated using Kolmogorov–Smirnov, Anderson–Darling, Chi-squared tests at critical value 0.05 for the four proposed gauge stations. Goodness of fit test results show that Gen. extreme value gives best results at Rampur, Sundergarh, Jondhra gauge stations followed by LP III whereas LP III best fits for Basantpur followed by Gen. extreme value.

Key words: confidence band, flow discharge, Gen. extreme value, Gumbel max, Log-Pearson III, Normal

Highlights

- Four statistical methods like Normal Distribution, Gumble Distribution, Log Pearson Type III and Extreme Distribution method are employed to forecast stream flow up to 150 years.
- Goodness fit test for above four statistical method also studied to find out the rank of data series at 5% significance level.
- Confidence band in the sense of maximum flow discharge is evaluated up to 95% of confidence limit.
- Sensitivity of all physical parameter also discussed on the four statistical methods.
- Hydrological data are discuss through various statistical indices.

INTRODUCTION

Consistent and precise stream flow forecasting are needed for innumerable issues such as water resources planning, strategy improvement, maneuver and upkeep events. In water management, forecasting high-quality stream flow and effective usage of this estimate gives substantial financial and communal assistance. For the hydrologic constituent, there is the requirement of interim as well as...
enduring events of stream flow forecasting for optimizing systems or for planning future growth or drop. Interim forecasting denotes hourly or day-to-day forecasting, which is vital for caution against flood and safety, and enduring forecasting is on the basis of monthly, seasonal or annual timescales which is very beneficial in reservoir processes and irrigation administration choices like distributing water to consumers downstream, arranging discharges, famine extenuation and handling river agreements or applying compacted acquiescence.

Masmoudi & Habaib (1993) developed seven statistical channeling 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. Evensen (1994) discussed a novel chronological data integration technique based on predicting error statistics utilizing Monte Carlo procedures which served as a superior alternative to solve customary and computationally enormous challenging estimated error covariance equation utilized in extended Kalman filter. Bartholmes & Todini (2005) studied the possibility to extend flood predicting lag times equal to 10 days by engaging an amalgamation of innovative climatological and hydrological models and presented outcomes of the joined approach among a numerical weather forecast system and rainfall-runoff model. Griffis & Stedinger (2007) explored features of LP III distribution in real and log space. Assessments with outlines of U.S. flood data revealed that LP III distribution offers a sensible model for yearly flood sequence distribution from unfettered catchments for log space skew. Moreover, for LP III distribution relations of L-moment ratio were established so as to compare them to overall statistics of a province. Rowinski et al. (2002) discussed two probability density functions, prevalent in hydrological studies, i.e., Log-Gumbel and Log-Logistic, with regard to use of the functions to hydrological numbers and problems ensuing from their mathematical properties. The maximum likelihood method promises merging of the estimators away from the area of reality of the two L-moments. Rath et al. (2018) employed the ARIMA model to predict flow discharge at Mahanadi river basin. Helsel & Hirsch (1992) discussed probabilistic approaches usually accomplished in hydrology. Gumbel max value and LP III distribution are considered to be the best prevalent probabilistic models related to solving water resources problems. Kamal et al. (2017) applied statistical distribution on discharge data for two locations and discovered that Log-normal is applicable for Haridwar and Gumbel EV1 for Garhmukteshwar. Subsequent to finding an appropriate distribution for a region, the distribution helps in predicting discharge for a certain return period. Brandimarte & Di Baldassarre (2012) proposed another method on the basis of applicability of uncertain flood profile to estimate uncertainty in hydraulic modeling and FFA, where the major considerable uncertainty sources are clearly scrutinized. Ewemoje & Ewemooje (2011) investigated Normal, Lognormal, and LP III distributions to model at-site annual peak flood flow in Ogun-Oshun River, Nigeria. Chen et al. (2011) analyzed risk of flooding resulting from occurrence of flood taking into consideration flood enormity and time of incidence applying LP III and mixed von Mises distribution. Mukherjee (2013) developed a mathematical model regarding peak flood discharge and return period utilizing GEV. Bezak et al. (2014) explored influence of threshold value in peaks-over-threshold method on FFA results and compared different statistical distribution functions and evaluated three parameter estimation techniques. Haddad & Rahman (2011a) investigated usability of quantile regression method as a feasible regional FFA technique for New South Wales, Australia. Haddad & Rahman (2011b) examined flood data from Tasmania, Australia considering an assortment of models’ criteria: Akaike Information Criterion (AIC), AIC-second order variant, Bayesian IC, and a customized ADC. Results obtained by simulating Monte Carlo model shows that ADC is better at recognizing parent allocation fittingly. Grimaldi & Serinaldi (2006) modeled trivariate joint distribution of flood peak, volume, and duration using a class of copulas called asymmetric Archimedean copulas. Hirabayashi et al. (2013) presented universal flood hazard for this century on the basis of results obtained from climate models and employed a condition of skill for universal stream steering model with a barrage system for computing river
discharge and flood area. Haddad & Rahman (2012) proposed a model utilizing Bayesian generalized least squares regression in an authoritative area structure for RFFA of ungauged watersheds in eastern Australia. Yue (2001) investigated usability of a two variable gamma model comprising five constraints to describe combined probability actions of multiple variable flood occurrences. Reis & Stedinger (2005) explored Bayesian Markov chain Monte Carlo techniques to evaluate subsequent circulation of flood magnitude, flood menace, and constraints of Log-normal and LP III distributions. Subyani (2011) quantified hydro-geological distinctiveness and probability of flood occurrence of several main valleys in western Saudi Arabia by applying GEV and LP III distributions to peak daily precipitation data. Sraj et al. (2015) examined 58 flood occurrences at Litija station on Sava River, Slovenia applying different bivariate copulas and contrasted them utilizing various arithmetic, graphic, and higher extremity reliance experiments. Merz & Thieken (2005) explored the difference between natural and epistemic uncertainty in FFA. Ouarda et al. (2001) projected an apparent theoretical framework for application of canonical correlations in RFFA using data of 106 stations from Ontario province, Canada. Micevski et al. (2015) presented a substitute RFFA technique that is predominantly valuable when adequately harmonized areas cannot be recognized on the basis of region of influence. Sahoo et al. (2020) studied bivariate low flow frequency analysis of Mahanadi basin, which has major deviations in hydrological performance from upstream to downstream, for two main low flow characteristics. Parhi (2018) estimated peak floods at Mahanadi River at the Hirakud dam and Naraj of up to 100 years’ recurrence interval utilizing HEC-RAS and Gumbel’s distribution. Pawar & Hire (2018) applied LP III distribution for flood data of four locations on the Mahi River and studied peak stream flow frequency, magnitude in field of flood hydrology. Lima et al. (2016) estimated local and regional GEV distribution for flood frequency analysis of Rio Doce basin, Brazil in a multilevel, hierarchical Bayesian framework, to explicitly model and reduce uncertainties. Bhat et al. (2019) carried out flood frequency analysis of the River Jhelum employing Gumbel and LP-III distributions for simulating future flood discharge scenarios from three positions. 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 to a rainfall-based flood frequency model and accounting probability of spatial and temporal rainfall pattern over the basin.

Here, various statistical methods are established for estimation of flow discharge at four gauge stations in Mahanadi river basin, India. Also, goodness of fit is applied for analyzing data sets. Flow discharge is calculated through various confidence limits (up to 95%) and is also discussed here.

**STUDY AREA**

Mahanadi (Figure 1) is one a major inter-state east flowing river in peninsular India. The river length from the origin to convergence in the Bay of Bengal is 851 km. In Chhattisgarh the river flows for 357 km and the other 494 km is in Odisha. Details of geographical and hydrological details of four gauging stations are shown in Table 1. Four gauge stations, Rampur, Sundargarh, Jondhra, and Basantpur are considered for our research.

**METHODOLOGY**

Generalized extreme value

Generalized extreme value is a continuous probability distribution developed within extreme value theory. It is a combination of Gumbel, Fréchet, and Weibull extreme value distributions and is a bounded distribution of standardized maxima of a series of autonomous and indistinguishable
dispersed arbitrary variables. GEV is utilized as an estimate for modeling maxima of lengthy (limited) series of arbitrary variables. Prominently, while using this distribution, the upper bound is unidentified and hence has to be projected; when Weibull is applied, the lower bound is identified as zero.

Frequency factor for GEV distribution is:

$$K_t = \frac{\sqrt{6}}{\pi} \left\{ 0.5772 + \ln \left[ \ln \left( \frac{T}{T - 1} \right) \right] \right\}$$  \hspace{1cm} (1)

To express $T$ in terms of $K_t$:

$$T = \frac{1}{1 - \exp \left\{ -\exp \left[ -\left( 0.5772 + \frac{\pi K_t}{\sqrt{6}} \right) \right] \right\}}$$  \hspace{1cm} (2)

**Figure 1** | Proposed river gauge stations.

**Table 1** | Details of geographical and hydrological data for gauge stations

| Hydro-meteorological station | Length of record (years) | Hydrological Mean | SD | Skewness | Kurtosis | Maximum flow discharge | Drainage area (km²) | Elevation from MSL (m) |
|-----------------------------|--------------------------|------------------|----|----------|----------|-----------------------|---------------------|-----------------------|
| Rampur                      | 29                       | 16,187.16        | 12,070.06 | 1.576    | 2.597    | 49,857.57             | 8,348.27            | 290                   |
| Sundargarh                  | 29                       | 36,514.85        | 13,998.42 | 1.276    | 1.637    | 74,916.31             | 9,183.73            | 243                   |
| Jondhra                     | 29                       | 92,300.91        | 52,456.69 | 1.432    | 2.526    | 242,549               | 10,930.43            | 272                   |
| Basantpur                   | 29                       | 225,305          | 110,452.4 | 1.533    | 3.472    | 561,700               | 10,672.87            | 236                   |
Predicted discharge ($Q_p$) is calculated with the standard normal distribution formula for the different return periods, and expressed as:

$$Q_p = \mu + K_t \sigma$$  \hspace{1cm} (3)

where $Q_p =$ predicted discharge, $\mu =$ standard mean, $\sigma =$ standard deviation.

**Normal distribution**

In statistics, normal distribution is a type where the data are characterized by a bell-shaped curve. Discrete form and curve location are obtained by mean and standard deviation. As many natural phenomena fit in this, it is a very significant probability distribution in statistics. This distribution illustrates how variable data are dispersed. The majority of annotations group about a central peak as it is symmetric and probability for data shrink off uniformly in both directions more away from mean. Arithmetic mean of sample $x_1, x_2, ... x_n$ typically represented by $\mu$ is the sum of the sampled value divided by item number(n):

$$\text{Simple mean (C)} = \frac{x_1 + x_2 + \ldots \ldots + x_n}{n} = \frac{1}{n} \sum_{i=1}^{n} x_n$$  \hspace{1cm} (4)

For the required return period ($T$), the probability factor ($P$) is evaluated in percentage. The conversion formula used to evaluate the probability is given as:

$$P = \frac{1}{T} (\%)$$  \hspace{1cm} (5)

From the standard normal distribution table, by interpolation, the frequency factor ($K_t$) is computed based on the different return periods, where frequency factor equals to standard normal deviate ($z$). Finally, the predicted discharge ($Q_p$) is found using the standard normal distribution formula for the different return periods for the respective seasons:

$$Q_p = \mu + K_t$$  \hspace{1cm} (6)

**Gumbel max**

Gumbel is a type of statistical distribution which began from extreme theory. Function in this distribution is unrestrained on whichever side, leading to negative flow calculation. This represents distribution of extreme values, either highest or lowest of samples, used in various distributions and for modeling distribution of peak levels. This is utilized for predicting earthquake, flood, and other natural hazards. It also models operational threat in managing threat and product life which wears out rapidly prior to a certain age. For the required return period ($T$), abridged variate ($Y_t$) has been assessed by using the formula:

$$Y_t = \ln(\ln (T/(T - 1)))$$  \hspace{1cm} (7)
The abridged mean and abridged standard deviation has been obtained from the Gumbel distribution table for the given sample size \(N\). Then the frequency factor is estimated using the formula:

\[
K_t = \frac{Y_t - Y_n}{S_n}
\]  

(8)

where \(K_t\) = frequency factor, \(Y_t\) = abridged variate, \(Y_n\) = abridged mean, \(S_n\) = abridged standard deviation.

Thus, the predicted discharge \(Q_p\) is computed using the standard normal distribution formula for diverse return period for respective seasons:

\[
Q_p = \mu + K_t\sigma
\]

(9)

Log-Pearson III

LP III is a method in statistics to fit frequency distribution values for predicting flood at a few sites of a specified river. Frequency distribution is built after calculating data related to statistics at a particular river site. Flood occurrence probability of different densities can be taken out from the curve. This particular method helps in extrapolating event data having return periods ahead of pragmatic occurrence of flood. After finding the actual discharge, then calculate the natural logarithm of the actual discharges \((Z)\) and find the standard logarithmic mean \((\mu)\) and standard logarithmic deviation \((\sigma)\) of the calculated discharges for the respective seasons:

\[
Z = \log_{10} Q
\]

(10)

Then the coefficient of skewness \((C_s)\) is calculated using the logarithmic discharges \((Z)\) and for the required return period \((T)\), we calculated the probability \((P)\) in percentage, as per the formula:

\[
P = \frac{1}{T} (\%)
\]

(11)

From the standard normal distribution table, by interpolation, we calculated the standard normal deviate \((z)\). The frequency factor depends on coefficient of skewness and return period. When \(Cs = 0\), the frequency factor is equal to standard normal deviate \(z\) and is calculated as in the case of Normal deviation. When \(Cs \neq 0\), the frequency factor \((K_t)\) is modified by using the formulae developed by Kite (1977):

\[
K_t = z + (z^2 - 1)k + \frac{1}{3}(z^3 - 6z)k^2 - (z^2 - 1)k^3 + zk^4 + \frac{1}{3}k^5
\]

(12)

where \(z\) = normal deviate

\[
k = \frac{Cs}{6}
\]

(13)

\(K_t\) = frequency factor

Now, predicted logarithmic discharge is calculated by using the formula:

\[
q_p = \mu + K_t\sigma
\]

(14)
where $q_p = \text{predicted logarithmic discharge}$, $\mu = \text{standard logarithmic mean}$, $\sigma = \text{standard logarithmic deviation}$.

Hence, the predicted discharge ($Q_p$) is calculated by taking the antilog of $q_p$:

$$Q_p = \text{antilog} (q_p) \text{ m}^3/\text{s}$$

**Goodness-of-fit test**

For a given set of data whether a certain distribution is fit or not is checked by 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 critical value for the constant significance level. Chi-squared, Anderson–Darling (AD) and Kolmogorov–Smirnov (KS) tests are employed here at significance level 0.05.

**Kolmogorov–Smirnov test**

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:

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

(15)

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

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

(16)

**Anderson–Darling test**

It associates fit of an observed to an expected CDF, hence giving additional weight to distribution tails than previous experiments.

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

(17)

**Chi-squared test**

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

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

(18)

where

$O_j = \text{observed frequency}$
expected frequency \( (E_j) = F(Y_2) - F(Y_1) \)

\( F \) = cumulative distribution function

\[ l = 1 + \log_2 m \]

where,

\( m \) = sample size.

**RESULTS AND DISCUSSION**

Parameters like shape \((k, \alpha)\), scale \((\sigma, \beta)\), location \((\mu, \gamma)\) for different distribution methods of the four gauge stations are presented in Table 2. Probability density function (PDF) and the cumulative density function (CDF) graph for respective gauge stations are displayed in Figure 2.

Three goodness-of-fit tests (as presented in the section ‘Goodness-of-fit test’) were used to analyze rainfall data series at the four stations chosen. Test statistics in correspondence to each test were calculated, and hypothesis testing was done at significance level 0.05. For KS, AD, and Chi-squared tests, the tests reject the hypothesis concerning distribution level if the statistics found are more than the critical value 2.5, 0.12555, and 12.592, respectively (Millington et al. 2011). KS, AD and Chi-squared tests were applied in Easy Fit software for selecting the best fit distribution \((s)\) and outcomes obtained are specified in Table 3.

**Table 2** | Details of distribution fitting parameters for GEV, LP III, Gumbel Max, and Normal method

| Sl. No. | Distribution | Parameters |
|---------|--------------|------------|
| Rampur  | Gen. extreme value | \( k = 0.21516, \sigma = 599.29, \mu = 842.82 \) |
| 1       | Gumbel max    | \( \sigma = 784.25, \mu = 896.25 \) |
| 2       | Log-Pearson III | \( \alpha = 9.4308, \beta = -0.257842, \gamma = 9.3736 \) |
| 3       | Normal        | \( \sigma = 1.0058, \mu = 1.3489 \) |
| Sundargarh | Gen. extreme value | \( k = 0.1665, \sigma = 766.83, \mu = 2450.5 \) |
| 1       | Gumbel max    | \( \sigma = 909.54, \mu = 2517.9 \) |
| 2       | Log-Pearson III | \( \alpha = 14.593, \beta = 0.09202, \gamma = 6.6161 \) |
| 3       | Normal        | \( \sigma = 1.1665, \mu = 3.0429 \) |
| Jondhra | Gen. extreme value | \( k = 0.1469, \sigma = 2891.3, \mu = 5355.6 \) |
| 1       | Gumbel max    | \( \sigma = 3408.4, \mu = 5724.4 \) |
| 2       | Log-Pearson III | \( \alpha = 157.42, \beta = -0.04439, \gamma = 15.793 \) |
| 3       | Normal        | \( \sigma = 4371.4, \mu = 7691.7 \) |
| Basantpur | Gen. extreme value | \( k = 0.1349, \sigma = 6117.9, \mu = 14310.0 \) |
| 1       | Gumbel max    | \( \sigma = 7176.6, \mu = 14633.0 \) |
| 2       | Log-Pearson III | \( \alpha = 2302.6, \beta = -0.00973, \gamma = 32135 \) |
| 3       | Normal        | \( \sigma = 9204.4, \mu = 18775.0 \) |
At Rampur, Sundergarh, and Jondhra gauge stations, extreme value distribution gives best results followed by LP III whereas LP III best fits for Basantpur followed by extreme value. Therefore, extreme value can be utilized to calculate flood return periods for the present study area. The poor ranking of Normal distribution fitted results is perhaps due to its nature. Given that Normal distribution is based on central limit theorem while the data considered in this study (annual maximum) are at the extreme right of all considered distributions, it was expected that normal fit to the data would be least efficient. In addition it is observed that at Rampur, Jondhra, and Basantpur stations the Chi-squared test correctly rejects normal fit to data as both statistics are related to central limit theorem.

**Gen. extreme value**

For Rampur watershed, the extreme value of flood calculated during monsoon period the range of peak flood lies within 177.4414 m$^3$/sec to 321.6385 m$^3$/sec for 10 years to 150 years’ return period (Table 4). Similarly for Sundargarh estimated flood fluctuates from 304.3543 m$^3$/sec to
471.5889 m³/sec. For Jondhra watershed, designed flood lies within 893.1144 m³/sec to 1,944.325 m³/sec for 10 years to 150 years return period. The magnitude of peak floods with respect to return period is found to be 2,052.522 m³/sec to 3,372.061 m³/sec for Basantpur watershed. This range is the highest among all seasonal peak floods.

**Gumbel max**

The intended flood value for Rampur watershed lies within 198.8535 m³/sec to 372.361 m³/sec for 10 years to 150 years' return period (Table 5). Correspondingly for Sundargarh, the appraised flood diverges from 329.1873 m³/sec to 530.415 m³/sec. For Jondhra watershed, premeditated flood lies within 986.1719 m³/sec to 2,133.888 m³/sec for 10 years to 150 years' return period. The magnitude of peak floods with respect to return period is found to be 2,248.463 m³/sec to 3,836.22 m³/sec for Basantpur watershed.
Normal method

For 10 years to 150 years’ return period the calculated flood value deviates within 175.76 m³/sec to 255.892 m³/sec for Rampur watershed (Table 6). Consistently for Sundargarh, the assessed flood is from 302.4046 m³/sec to 395.3386 m³/sec. For Jondhra watershed, premeditated flood contrasts within 885.808 m³/sec to 1,234.06 m³/sec for 10 years to 150 years’ return period. The enormousness of extreme flood with respect to return period is found to be 2,037.138 m³/sec to 2,770.419 m³/sec for Basantpur watershed.

Log-Pearson III

The gauged flood value diverges within 177.4024 m³/sec to 317.6723 m³/sec for 10 years to 150 years’ return period for Rampur watershed (Table 7). Reliably for Sundargarh, the projected flood is from 303.5037 m³/sec to 532.3849 m³/sec. For Jondhra watershed, the planned flood contrasts within
897.3183 m$^3$/sec to 2,183.191 m$^3$/sec for 10 years to 150 years' return period. The enormousness of extreme flood with respect to return period is established to be 2,047.682 m$^3$/sec to 3,525.389 m$^3$/sec for Basantpur watershed.

Actual data from 2011 to 2019 are considered here for testing purposes. Comparison graphs of observed and simulated flood discharge for all proposed stations are presented in Figure 3.

Confidence band for difference scenario

For a given return period, $x_T$ is determined by Gumbel methods which have errors because of limited use of sample data. The confidence interval indicates the limits about the calculated value between which the true value can be said to lie with a specific probability based on sampling errors only. Confidence interval of variate is bounded by value $x_1$, $x_2$ for a confidence probability $c$ is $x_2 = x_1 \pm f(c)S_e$:
where \( f(c) \) function of confidence probability is:

\[
C(\% \ f(c)) = 0.0674 \times 1.645 + 0.5
\]

\[
S_e = \text{probable error} = b \frac{a_n - 1}{\sqrt{N}}
\]
Table 5 | Flow discharge with respect to return period at four gauge stations

| Return period (year) | Discharge (m³/sec) | Rampur | Sundargarh | Jondhra | Basantpur |
|----------------------|---------------------|--------|------------|---------|-----------|
| 10                   | 198.8535            | 329.1873 | 986.1719  | 2,248.463 |
| 20                   | 244.2809            | 381.8724 | 1,183.6    | 2,664.167 |
| 30                   | 270.1493            | 411.8736 | 1,296.025  | 2,900.887 |
| 35                   | 280.2443            | 423.5814 | 1,339.898  | 2,993.265 |
| 40                   | 288.4465            | 433.094  | 1,375.544  | 3,068.323 |
| 50                   | 302.958             | 449.9239 | 1,438.612  | 3,201.117 |
| 60                   | 314.3149            | 463.0952 | 1,487.969  | 3,305.043 |
| 70                   | 324.4099            | 474.803  | 1,531.842  | 3,397.422 |
| 75                   | 328.8264            | 479.9251 | 1,551.036  | 3,437.837 |
| 100                  | 347.1236            | 501.1455 | 1,842.462  | 3,605.273 |
| 150                  | 372.361             | 530.415  | 2,133.888  | 3,856.22  |

Table 6 | Flow discharge with respect to return period at four gauge stations

| Return period (year) | Discharge (m³/sec) | Rampur | Sundargarh | Jondhra | Basantpur |
|----------------------|---------------------|--------|------------|---------|-----------|
| 10                   | 175.76              | 302.4046 | 885.808   | 2,037.138 |
| 20                   | 200.571             | 331.1791 | 993.636   | 2,264.179 |
| 30                   | 213.312             | 345.9552 | 1,049.01  | 2,380.768 |
| 35                   | 217.335             | 350.6214 | 1,066.49  | 2,417.585 |
| 40                   | 221.358             | 355.2875 | 1,083.98  | 2,454.403 |
| 50                   | 227.393             | 362.2867 | 1,110.21  | 2,509.629 |
| 60                   | 232.758             | 368.5082 | 1,133.52  | 2,558.719 |
| 70                   | 236.781             | 373.1744 | 1,151.01  | 2,595.536 |
| 75                   | 238.793             | 375.5075 | 1,159.75  | 2,613.945 |
| 100                  | 248.985             | 387.3283 | 1,204.05  | 2,707.216 |
| 150                  | 255.892             | 395.3386 | 1,234.06  | 2,770.419 |

Table 7 | Flow discharge with respect to return period at four gauge stations

| Return period (year) | Discharge (m³/sec) | Rampur | Sundargarh | Jondhra | Basantpur |
|----------------------|---------------------|--------|------------|---------|-----------|
| 10                   | 177.4024            | 303.5037 | 897.3183  | 2,047.682 |
| 20                   | 216.9286            | 357.3192 | 1,038.848 | 2,425.163 |
| 30                   | 238.7493            | 389.9993 | 1,144.289 | 2,644.386 |
| 35                   | 245.8349            | 401.1399 | 1,192.41  | 2,717.519 |
| 40                   | 253.0086            | 412.7043 | 1,215.78  | 2,792.611 |
| 50                   | 265.9266            | 430.8866 | 1,177.567 | 2,909.031 |
| 60                   | 273.7815            | 447.9389 | 1,367.25  | 3,016.452 |
| 70                   | 281.2598            | 461.3095 | 1,500.813 | 3,099.53  |
| 75                   | 285.0255            | 468.1896 | 1,531.547 | 3,141.897 |
| 100                  | 304.3574            | 505.1676 | 1,818.622 | 3,365.321 |
| 150                  | 317.6723            | 532.5849 | 2,183.191 | 3,525.389 |
where,

\[ K = \text{frequency factor given by } \frac{y_T - y_n}{S_n} \]

\[ \sigma_n - 1 = \text{standard deviation} \]

\[ N = \text{sample size} \]

\[ b = \sqrt{1 + 1.3k + 1.1K^2} \]
For different values of $T$, $X_T$ is calculated and shown in Figure 4. Also 95, 90, 85, 80, and 75% confidence limits for various values of $T$ are shown. It is seen that while confidence probability rises, confidence interval also increases. Further increase in $T$ causes the confidence band to spread. Thus, Gumbel distribution will give erroneous results if the sample has a value of $C_s$ very much different from 1.14.

**Sensitivity analysis**

For Normal distribution method, probability factor is dependent on required return period ($T$), which is inversely proportional. Frequency factor ($K_t$) varies with return periods. Predicted discharge ($Q_p$)
increases with respect to the increase in required return period, while probability factor \((P)\) decreases. When frequency factor increases simultaneously predicted discharge increases. Predicted flood increases with regard to the increase in the required return period, at the same time, frequency factor increases with decrease of standard deviation in the case of Gen. extreme value method. Predicted flood increases with reference to the increase in the required return period, at the same time, frequency factor also increases, whereas reduced mean \((Y_n)\) and reduced standard deviation \((S_n)\) remain constant for all recurrence intervals; however, reduced variate \((Y_t)\) increases by Gumbel max. In LP III, predicted flood increases with increase in the required return period, at the same time, frequency factor also increases, whereas coefficient of skewness \((C_s)\) and reduced standard deviation remain constant for all recurrence intervals.

**CONCLUSIONS**

In this paper, an effort has been made to forecast discharges at various return periods using statistical methods. Here, four statistical methods are used to predict flow discharge in the Mahanadi river basin, covering four stations. Four statistical distribution methods, namely, Normal, LP III, Gumbel max, and Gen. extreme value method are employed here. Based on the trends of the last 60 years, the maximum and minimum discharges are found at 150 years and 10 years’ return period, respectively. The rate of increase of discharge is very high at the initial return periods and then it becomes constant and eventually lower. The shapes of the graphs are common in nature and most of the time they do not intersect with each other. In most of the cases, Gumbel max gives the peak flood

![Figure 4](https://iwaponline.com/h2open/article-pdf/doi/10.2166/h2oj.2020.004/705688/h2oj2020004.pdf)
discharge and normal distribution contributes to the least discharge. The Gumbel max is the most widely used method to obtain flood discharge as this can be used for infinite sample sizes. The influencing factor of frequency is analyzed on the basis of analysis of the 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 modeling and development for water resources engineering.

**DATA AVAILABILITY STATEMENT**

All relevant data are included in the paper or its Supplementary Information.

**REFERENCES**

Bartholmes, J. & Todini, E. 2005 Coupling meteorological and hydrological models for flood forecasting. *Hydrology and Earth System Sciences Discussions* 9 (4), 333–346.

Bezak, N., Brilly, M. & Šraj, M. 2014 Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. *Hydrological Sciences Journal* 59 (5), 959–977.

Bhat, M. S., Alam, A., Ahmad, B., Kothia, B. S., Farooq, H., Taloor, A. K. & Ahmad, S. 2019 Flood frequency analysis of Jhelum in Kashmir basin. *Quaternary International* 507, 288–294.

Brandimarte, L. & Di Baldassarre, G. 2012 Uncertainty in design flood profiles derived by hydraulic modelling. *Hydrology Research* 43 (6), 753–761.

Chen, L., Singh, V. P., Shenglan, G., Hao, Z. & Li, T. 2011 Flood coincidence risk analysis using multivariate copula functions. *Journal of Hydrologic Engineering* 17 (6), 742–755.

Evensen, G. 1994 Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans* 99 (C5), 10143–10162.

Ewemoje, T. A. & Ewemooje, O. S. 2011 Best distribution and plotting positions of daily maximum flood estimation in Ona River in Ogun-Oshun river basin. *Nigeria. Agricultural Engineering International: CIGR Journal* 13 (5), 1–11.

Griffies, S. 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.

Grimaldi, S. & Serinaldi, F. 2006 Asymmetric copula in multivariate flood frequency analysis. *Advances in Water Resources* 29 (8), 1155–1167.

Haddad, K. & Rahman, A. 2011a Regional flood estimation in New South Wales Australia using generalized least squares quantile regression. *Journal of Hydrologic Engineering* 16 (11), 920–925.

Haddad, K. & Rahman, A. 2011b Selection of the best fit flood frequency distribution and parameter estimation procedure: a case study for Tasmania in Australia. *Stochastic Environmental Research and Risk Assessment* 25 (3), 415–428.

Haddad, K. & Rahman, A. 2012 Regional flood frequency analysis in eastern Australia: Bayesian GLS regression-based methods within fixed region and ROI framework–Quantile Regression vs. Parameter Regression Technique. *Journal of Hydrology* 430–431, 142–161.

Helsel, D. R. & Hirsch, R. M. 1992 *Statistical Methods in Water Resources*, 1st edn, Studies in Environmental Science, Vol. 49. Elsevier.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H. & Kanae, S. 2013 Global flood risk under climate change. *Nature Climate Change* 3 (9), 816–821.

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.

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.

Masmoudi, M. & Habaieb, H. 1995 The performance of some real-time statistical flood forecasting models seen through multicriteria analysis. *Water Resources Management* 7 (1), 57–67.

Merz, B. & Thieken, A. H. 2005 Separating natural and epistemic uncertainty in flood frequency analysis. *Journal of Hydrology* 309 (1–4), 114–132.

Micevski, T., Hackelbusch, A., Haddad, K., Kuczera, G. & Rahman, A. 2015 Regionalisation of the parameters of the log-Pearson 3 distribution: a case study for New South Wales, Australia. *Hydrological Processes* 29 (2), 250–260.

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.
Ouarda, T. B., Girard, C., Cavadias, G. S. & Bobée, B. 2001 Regional flood frequency estimation with canonical correlation analysis. *Journal of Hydrology* 254 (1–4), 157–173.

Parhi, P. K. 2018 Flood management in Mahanadi Basin using HEC-RAS and Gumbel's extreme value distribution. *Journal of The Institution of Engineers (India): Series A* 99 (4), 751–755.

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.

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.

Reis Jr., D. S. & Stedinger, J. R. 2005 Bayesian MCMC flood frequency analysis with historical information. *Journal of Hydrology* 313 (1–2), 97–116.

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: 1. *Hydrological Sciences Journal* 47 (1), 107–122.

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.

Sraj, M., Bezak, N. & Brilly, M. 2015 Bivariate flood frequency analysis using the copula function: a case study of the Litiča station on the Sava River. *Hydrological Processes* 29 (2), 225–238.

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.

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.

Yue, S. 2001 A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrological Processes* 15 (6), 1033–1045.