Open access e-Journal
**Earth Science India** eISSN: 0974 – 8350
14 (3), 2021, 122-131
http://www.earthscienceindia.info/

Crossref

OPEN ACCESS

# Application of Kumaraswamy Distribution for Maximum Flood Heights at Naraj Barrage in Mahanadi River Basin, Cuttack, Orissa

**Nagesh S. and Laxmi Dharmannavar**
Department of Statistics, Karnatak University Dharwad, Karnataka-580003, India
Email: nageshstatistics@gmail.com; laxmi.d2420@gmail.com

## ABSTRACT

Naraj barrage with 1302m length and 6.9m height across Kathajodi River is located 14 km west of Cuttack town in Orissa. It is constructed to control flooding in Mahanadi and Kathajodi Rivers. During the past two decades Naraj barrage experienced several high floods in excess of 35,000 $m^3/s$ in the years 2001, 2003, 2006, 2008, 2011 and 2018. Kumaraswamy distribution unfortunately not yet been used by statisticians due to less familiarity of the distribution despite being a good number of applications in hydrology. In this work, Kumaraswamy distribution is identified as one of the most suitable alternative flood frequency distribution for maximum flood heights and their return periods at Naraj barrage in Mahanadi River Basin (MRB) by considering commonly used ten candidate distributions along with this distribution. Different techniques of parametric estimation were used to obtain parameters value of the candidate distributions. Performed goodness of fit tests of Kolmogorov–Smirnov and Anderson-Darling for obtaining best fitted distributions, Akaike Information Criteria, Bayesian Information Criteria and graphical techniques (PDF, CDF, Probability Difference, P-P and Q-Q plots) were used for comparing the candidate distributions. Also conducted simulation studies in R Programming for checking whether distribution imitate the observed values of flood heights. High quantiles of extreme flood heights are obtained for their return period.

**Keywords:** Naraj barrage; Kumaraswamy distribution; Goodness of fit test; MLE; Return period

## INTRODUCTION

The deleterious floods are the most dangerous natural disasters. It results in huge loss to human society by destroying crops, lives and properties in the affected regions. India is more susceptible to floods due to its geographical structure, in fact developing countries are more vulnerable to natural disasters than developed countries (Nagesh and Dharmannavar, 2021). Orissa and west Bengal witnessed number of cyclones and hurricanes in recent past due depression in Bay of Bengal. This results in heavy rainfall causing flood in major or other tributaries draining Orissa and adjoining areas. Flood generally occurs as a result of the large run-off produced from the river catchment by heavy rainfall. As the flow in river exceeds the river flow carrying capacity, the excess flow overtop the banks of river and spread over the flood plain areas especially in low lying regions (Solomon and Prince, 2013).

Please cite this article as: *Nagesh, S. and Dharmannavar, Laxmi (2021) Application of Kumaraswamy Distribution for Maximum Flood Heights at Naraj barrage in Mahanadi river basin, Cuttack, Orissa. Earth Science India, v. 14 (3), pp. 122-131 https://doi.org/10.31870/ESI.14.3.2021.7*

Reliable estimation of maximum flood discharge (MFD) or peak discharge at a specific return period is necessary for proper planning and design of hydraulic structures like dams, spillways, culverts etc. As the hydrologic phenomena governing the MFD of annual maximum discharge is highly stochastic in nature, the MFD can be effectively simulated using the best fitted probability distributions function (PDFs) to the series of recorded annual maximum discharge (Vivekanandan, 2015).

The river Mahanadi is one of the important inter-state east flowing rivers in peninsular India. It originates at the Amarkantak Hills of the Bastar plateau near Pharasiya village in Raipur district of Chattisgarh, India, it traverses with its 14 tributaries about 850 km and the total catchment area of the river basin is 1,41,600 sqkm (Guru and Jha, 2016). It is also called sorrow of Orissa because of natural disasters such as devastating floods are frequent visitors in this area. The deltaic region is faced the problems of flood, drainage and salinity more or less in every year due to occurrence of low level escapes. These escapes began at 17,000 cumsecs of undivided flood at Naraj site and made major challenges to the whole part of the deltaic area. Also recommended to provide non-structural measures along with structural measures to mitigate flood effects (Khatua and Patra, 2004).

In hydrological studies, the primary problem is to choose a frequency distribution function that fits the extreme flood observations in a particular site or region (Karim and Chowdhury, 1995). Several studies have been done for improving probability frequency distribution models to delineate the frequency distribution of extreme flood. Smithers (2012) remarked for necessity to improve the current methods and assess the new methods by other developed countries in the interest of which may be suitable for the basin and catchments of the developing countries for flood design estimates. Guru and Jha (2014) studied seven probability distributions to choose appropriate one for nineteen sites of the MRB by using L-Moment Method. Sukla *et al*. (2014) explored Seventeen probability distributions to select best one for four sites of Mahanadi Delta region through use of Kolmogorov-Smirnov and Anderson-Darling goodness of fit tests. Krishna and Veerendra (2015) analysed the flood frequency and magnitude of maximum monthly data by applying Gumbel distribution at Prakasham barrage. Naz *et al.* (2019) carried out frequency analysis at Guddu barrage of Indus River. Nevertheless, crucial problems come out when selecting the best distribution to describe analysis of extreme hydrological events, since there is no general acceptance as to which method or distributions should be used (Singo *et al*., 2012). Information related to this can be found in the literature (Nayak and mandal, 2019; Pai-Panandiker *et al.,* 2020; Sharma *et al.,* 2020; Swethapadma and Ojha, 2017). Kumaraswamy (1980) introduced Double Bounded Probability Distribution Function (DB PDF) also called Kumaraswamy distribution with an application to daily rainfall data, Sundar and Subbaiah (1989) applied this probability density function to describe ocean wave height. Though it has a good number of applications in hydrology, unfortunately not yet used widely by the statisticians due to less familiarity of the distribution.

In the present work, Kumaraswamy distribution is identified as one of the most suitable flood frequency distributions for maximum flood heights and their return periods at Naraj barrage by investigating eleven candidate distributions including two-parameters (2P); Gamma (Gamma2P), lognormal (Lognormal2P), Weibull (Weibull2P), three-parameters (3P); Gamma (Gamma3P), lognormal (Lognormal3P), Weibull (Weibull3P), Log-Pearson type 3(LP3), Generalized Extreme Value (GEV), Generalized Gamma (GG), and Gumbel distribution.

## MATERIALS AND METHODS

Forty-Six years (1972-2017) daily water level (in metres) data recorded thrice a day at Naraj barrage were collected from Central Water Commission (CWC), Bhubaneswar, which is authority for water management of MRB. Raw data made ready data by considering the procedures of data cleaning, data processing and filling of missing values for analysis of extreme flood heights.

Flood Frequency analysis is performed at Naraj barrage of MRB using Block Maxima (at-site) method. In this method hydrological years are considered as independent and identically distributed blocks for this study. Best three flood frequency distributions were obtained by considering Kumaraswamy distribution along with commonly used ten candidate distributions in hydrology.

The distribution function and probability density function of Kumaraswamy distribution is as follows.

$$F(z,\alpha,\beta,c,d) = 1 - \left[1 - \left[\frac{z-c}{d-c}\right]^{\alpha}\right]^{\beta}; \qquad z \in (c,d)$$

$$f(z,\alpha,\beta,c,d) = \frac{1}{(d-c)}\alpha\beta\left[\frac{z-c}{d-c}\right]^{\alpha-1}\left[1 - \left[\frac{z-c}{d-c}\right]^{\alpha}\right]^{\beta-1}; \qquad z \in (c,d)$$

where $\alpha > 0$ and $\beta > 0$ are the shape parameters and c and d are boundary parameters.

Different techniques of parametric estimation were used to obtain parameters values of the candidate distributions. Performed goodness of fit tests of Kolmogorov–Smirnov and Anderson-Darling for obtaining best fitted distributions, Akaike Information Criteria (Akaike, 1998) and Bayesian Information Criteria (Langat *et al.,* 2019), and graphical techniques (PDF, CDF, Probability Difference, P-P and Q-Q plots) were used for comparing the candidate distributions. Also conducted simulation studies in R Programming for checking whether distribution imitate the observed values of flood heights. High quantiles of extreme flood heights are obtained for their return period.

The methods used for the river water level data analysis by making use of R-programming software packages. Simulation study is also done to check whether the distribution imitate the observed values.

## RESULTS AND DISCUSSIONS

The results obtained by descriptive statistics of observed flood height data at Naraj barrage show negatively skewed distribution. The results of parameter estimation, goodness of fit tests and expected return periods are as follows. Fig. 1 depicts the Time series plot of Naraj site.

### Estimation of Parameters:

Parametric estimation techniques such as Maximum Likelihood Estimation (MLE) for the distributions Gamma3P, Generalised Gamma(GG), Kumaraswamy, Lognormal2P, Lognormal3P, and Weibull3P; Method of Moments for Gamma2P and Log-Pearson type 3; L-Moments for GEV and Gumbel; Least Square methods for Weibull2P were used to estimate the parameters of the eleven candidate distributions at Naraj barrage. For the preceding parametric estimation methods, the R Programming software packages were utilised to estimate parameters. Table-1 shows the estimated parameters.

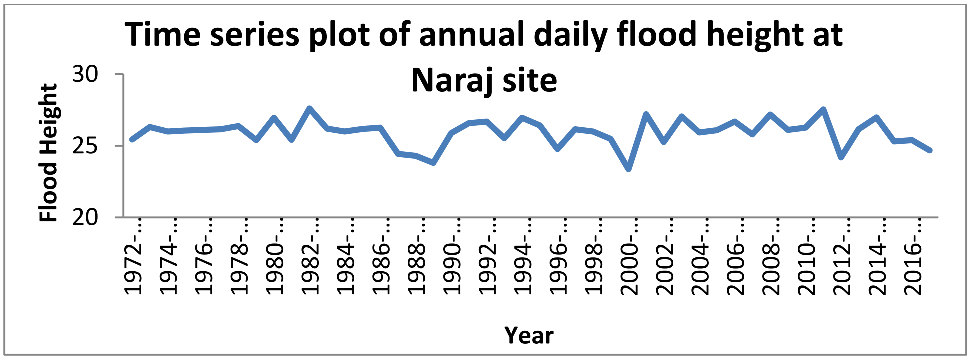


**Fig. 1:** Time series plot of annual daily flood height at Naraj barrage.

**Table-1:** Estimated parameters of the candidate distributions

| Distribution | Estimated Parameters |
|---|---|
| Weibull3P | $\alpha$=10.484, $\beta$=8.4617, $\gamma$=17.862 |
| Kumaraswamy | $\alpha$=7.2042, $\beta$=9.0722, c=20.022, d=28.656 |
| Weibull2P | $\alpha$=32.029, $\beta$=26.319 |
| Log-Pearson type 3 | $\alpha$=6.4197, $\beta$=-0.01457, $\gamma$=3.3481 |
| GEV | k=-0.5343, $\sigma$=1.0175, $\mu$=25.714 |
| Gamma2P | $\alpha$=754.31, $\beta$=0.03437 |
| Lognormal2P | $\sigma$=0.0365, $\mu$=3.2546 |
| Lognormal3P | $\sigma$=0.0242, $\mu$=3.6618, $\gamma$=-13.023 |
| Gamma3P | $\alpha$=226.05, $\beta$=0.0642, $\gamma$=11.416 |
| Gumbel | $\sigma$=0.736, $\mu$=25.502 |
| Gen. Gamma | k=1.2125, $\alpha$=674.16, $\beta$=0.03437 |

### Goodness of Fit Analysis:

Anderson-Darling (AD) and Kolmogorav-Smirnov (KS) goodness of fit tests analysis were performed to the observed data for verifying goodness of fit quality. Ranks were assigned to the candidate distributions based on p-value or test statistic of KS and AD independently with the distribution which has significant p-value or test statistic occupied 1$^{st}$ rank among eleven candidate distributions, the distribution which has second significant p-value occupied 2$^{nd}$ rank and so on. Total the ranks obtained by KS and AD statistic for each distribution independently and the smallest rank distribution of it indicates the best fit to the data. Lower the rank better the fit, that is rank 1 indicates best fit, rank 2 indicates second best fit and so on. Model evaluation criteria; Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) were also calculated for identifying the best distribution based on the distribution which has smallest value of AIC and BIC selected as best fitted distribution, second smallest value of AIC and BIC indicates second best fitted frequency distribution and so on. AIC and BIC criteria are most suitable for two parameters distributions than three or more (Haddad and Rahman, 2011). The best three frequency distributions obtained were Weibull3P, Kumaraswamy and Log-Pearson type 3 distribution. The ranks for each distribution, AIC, BIC and the mean correlation coefficient of the 30 simulated samples are given in the Table-2.

**Table-2:** Fitted distribution's Rank, AIC, BIC and the mean correlation coefficient of the 30 simulated Samples.

| **Distribution** | **KS** | **AD** | **Total rank** | **AIC** | **BIC** | **Correlation Coefficient** | **Remark** |
|---|---|---|---|---|---|---|---|
| Weibull3P | 1 | 1 | 2 | 126.5097 | 131.9956 | 0.9623 | Best model |
| Kumaraswamy | 2 | 2 | 4 | 128.4788 | 135.7933 | 0.9555 | 2nd best model |
| Weibull2P | 3 | 4 | 7 | 124.8837 | 128.5410 | 0.9256 | Good model |
| Log-Pearson type 3 | 4 | 3 | 7 | 126.5681 | 132.0540 | 0.9520 | 3rd best model |
| GEV | 5 | 5 | 10 | 130.5966 | 136.0826 | 0.9245 | Good model |
| Gamma2P | 6 | 6 | 12 | 129.0060 | 132.6633 | 0.9173 | Bad model |
| Lognormal2P | 7 | 8 | 15 | 129.4227 | 133.0800 | 0.9041 | Bad model |
| Lognormal3P | 8 | 7 | 15 | 131.0063 | 136.4922 | 0.8922 | Bad model |
| Gamma3P | 9 | 9 | 18 | 206.4817 | 209.2841 | 0.8900 | Bad model |
| Gumbel | - | - | - | 166.8900 | 170.5473 | 0.8047 | Not suitable |
| Gen. Gamma | - | - | - | 294.3264 | 297.8948 | 0.7899 | Not suitable |

Goodness of fit quality of fitted distributions can be checked by employing numerous graphical functions such as PDFs, CDFs, P-P, Q-Q, and PD plots. Fig. 2 and Fig. 3 depict the PDFs plots of the eleven candidate distributions and best three distributions respectively. These plots show Weibull3P, Kumaraswamy, and Log-Pearson type 3 distributions are good fit to the data. Fig. 4 and Fig. 5 depict the P-P plots and probability difference plots for the best three distributions. All the plotted points are reasonably close to the line of best fit and the three distributions show excellent results in modelling the upper tail. Generally, all the three distributions have probability differences of almost 0.00 for the upper tail and within 0.04 for the lower tail, but within 0.08 near the centre of the data which indicates good models. Q-Q plots were constructed in Fig. 6 and obtained similar results as in P-P plots.

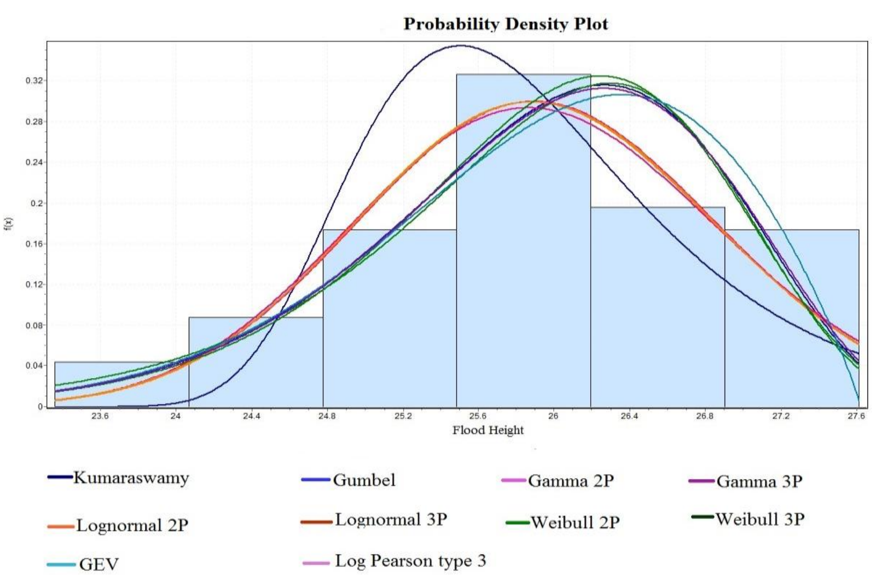


**Fig. 2:** Probability Density Plot containing 11 distributions.

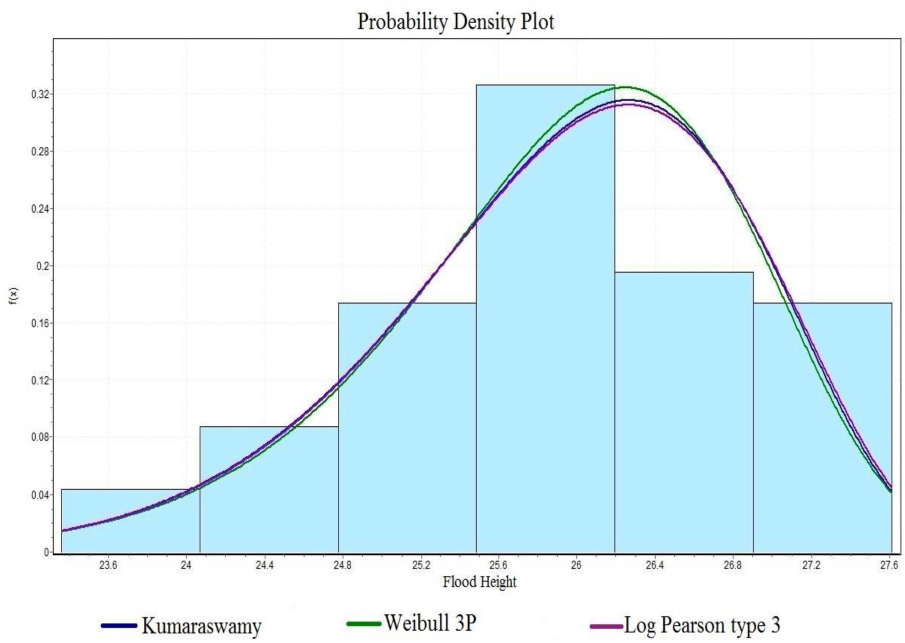


**Fig. 3:** Probability Density Plot of best 3 distributions.

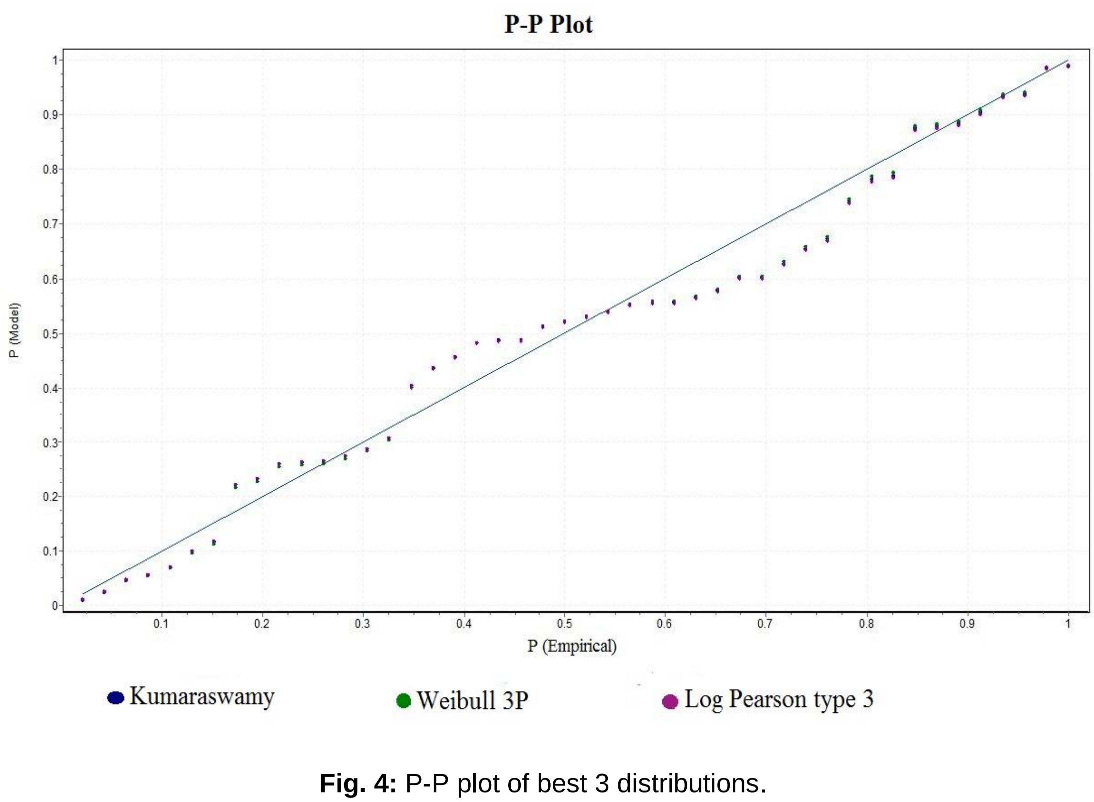


**Fig. 4:** P-P plot of best 3 distributions.

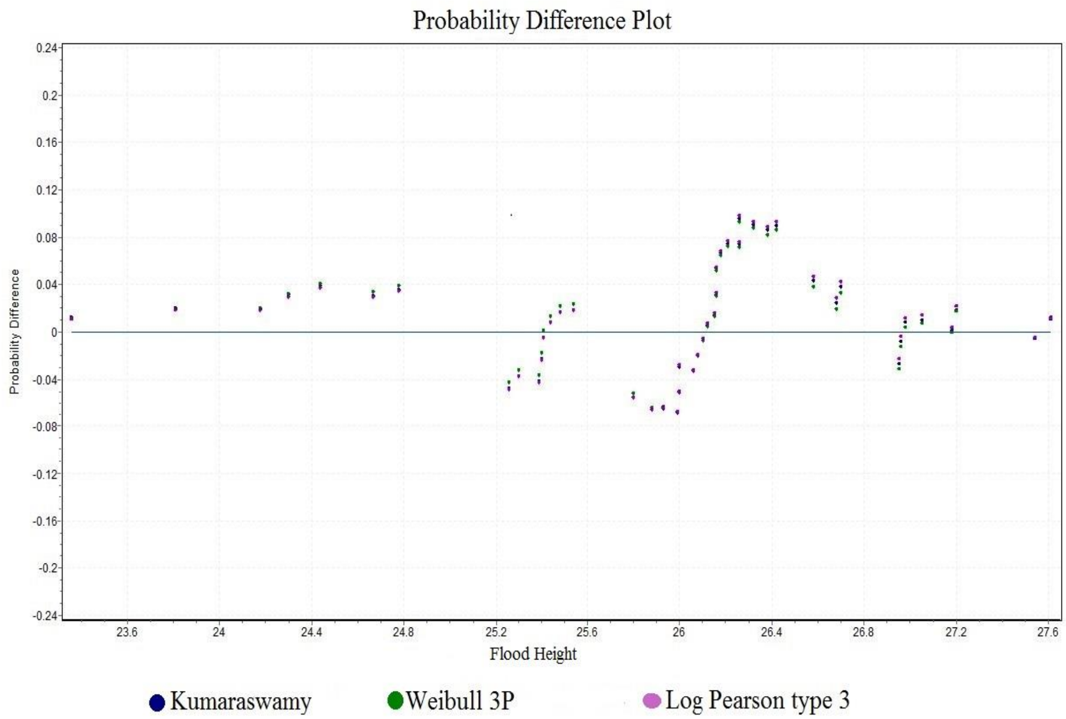


**Fig. 5:** P-D plot of best 3 distributions.

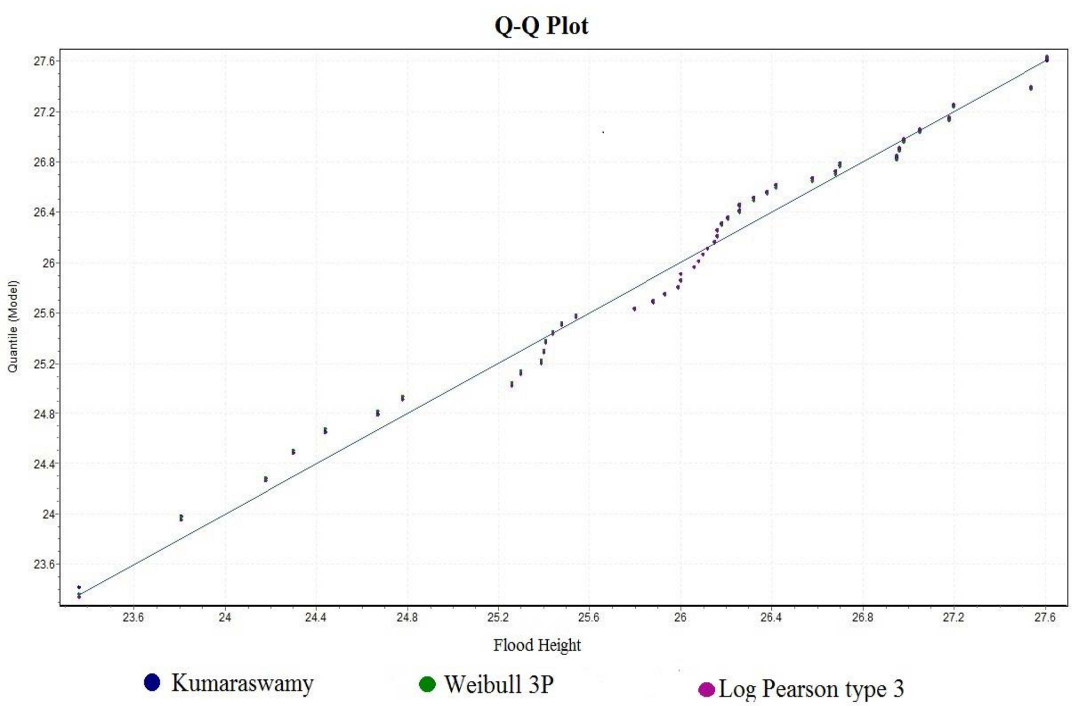


**Fig. 6:** Q-Q plot of best 3 distributions.

**Return Level Estimation:**

Fig. 7 shows graphical results of CDFs plots of the best three distributions at Naraj barrage. The CDF graphs show a single line for the site and the graph show the non-exceedance probabilities of the desired flood heights which are very useful to calculate their return periods at the site. The maximum flood heights return period is obtained by estimating high quantile results for best three distributions at the site as shown in Table-3. Estimated flood height and corresponding return period are very useful to water engineers in taking necessary steps to avoid bad effects of severe floods. It is also important to hydrometric scientists for forecasting flood height, formulating economic policies and agricultural planning (Nagesh and Dharmannavar, 2021). It is understood that Kumaraswamy distribution can be applied as an alternative distribution to traditional extreme value distribution in barrage flood frequency analysis.

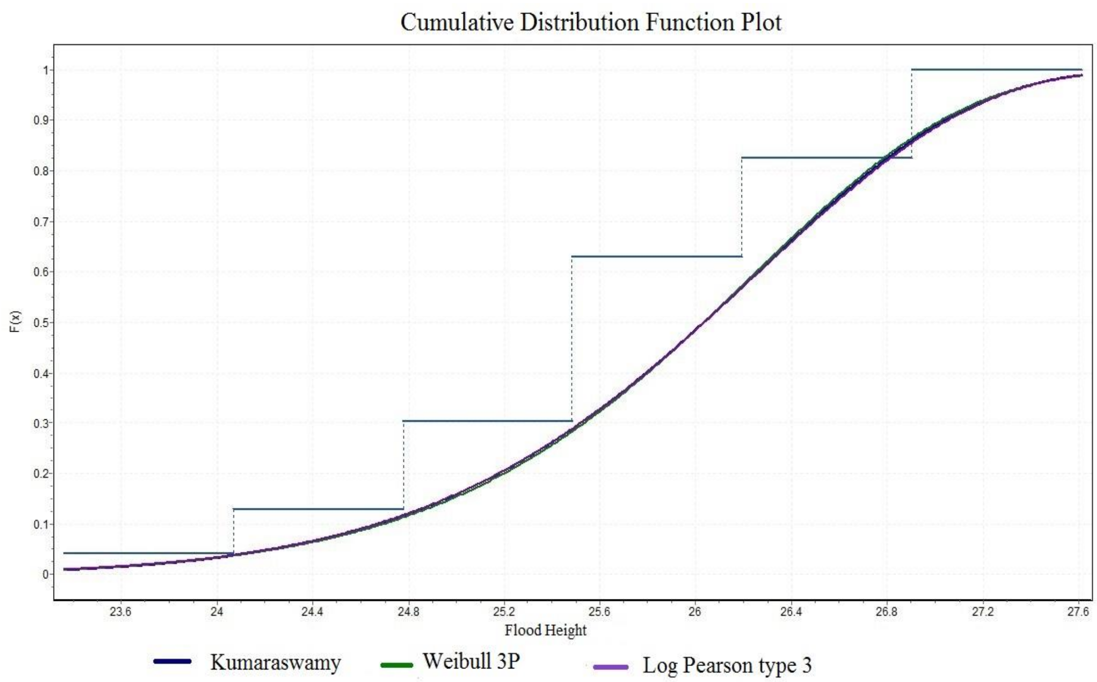


**Fig. 7:** CDF plot of best 3 distributions.

## CONCLUSIONS

Investigating appropriate flood frequency distributions still remains an active research area in hydrology. In this study, we obtained Weibull3P, Kumaraswamy and Log-Pearson type 3 distributions the best three distributions and suitable for modelling frequency analysis of floods at Naraj site in MRB by testing eleven candidate distributions for their good fit. Both the results of analytical and graphical functions analysis carried out by R Programming software packages indicated that Weibull3P, Kumaraswamy and Log-Pearson type 3 distributions as 1$^{st}$, 2$^{nd}$ and 3$^{rd}$ best distributions respectively and also observed that three parameter distributions more suitable for modeling the flood frequency analysis. Due to less familiarity of Kumaraswamy distribution to the statisticians, it has not been used for frequency analysis in the MRB. It is shown that Kumaraswamy distribution is one of the best distributions for maximum flood height and its frequency in hydrological applications. The maximum flood height 27.61 metres appeared in the year 1983. It is expected return periods in excess of 100 years or approximately 200 years. Though MRB has 40 Gauge and 25 Gauge and Discharge hydrometric stations, Naraj barrage contribute lot to flood in the deltaic region of MRB. Effect of floods can be minimized by avoiding deforestation, encroachment of river bank and urbanization near river bank in addition to structural and non-structural measures.

**Table-3:** Expected return levels of best 3 distributions.

| Expected return period **T**(years) | Exceedance probability **P** | Non-exceedance probability or CDF F(x) | Weibull3P | Kumara-swamy | Log-Pearson type 3 |
|---|---|---|---|---|---|
| 2 | 0.5 | 0.5 | 7.033 | 7.032 | 7.033 |
| 5 | 0.2 | 0.8 | 7.717 | 7.732 | 7.74 |
| 10 | 0.1 | 0.9 | 8.024 | 8.038 | 8.05 |
| 25 | 0.04 | 0.96 | 8.322 | 8.323 | 8.335 |
| 50 | 0.02 | 0.98 | 8.499 | 8.486 | 8.497 |
| 100 | 0.01 | 0.99 | 8.651 | 8.62 | 8.628 |
| 200 | 0.005 | 0.995 | 8.782 | 8.732 | 8.735 |
| 250 | 0.004 | 0.996 | 8.821 | 8.764 | 8.766 |
| 500 | 0.002 | 0.998 | 8.934 | 8.855 | 8.852 |
| 1000 | 0.001 | 0.999 | 9.037 | 8.934 | 8.925 |

It is interesting that most of the studies on flood frequency analysis used discharge data in cubic metres per second, but no studies of flood frequency analysis have used water level or flood height data in the MRB. The study used flood height data for obtaining the results. These results are very helpful to water practitioners for taking necessary measures in avoiding the bad effects of flooding.

**Conflict of Interest:** The authors declare that no conflict of interest exists regarding publication of the paper.

***Acknowledgement:*** The authors thank the Central Water Commission (CWC), Bhuvaneshwar, the authority for water resource management in India under the Ministry of Jal Shakti for providing data, the authors also thank to the Reviewer and the Editor for their comments and suggestions.

## REFERENCES

Akaike, H. (1998) Information theory and an extension of the maximum likelihood principle. *In:* Selected papers of Hirotugu Akaike, Springer, New York, NY, pp. 199-213.

Guru, Nibedita and Jha, Ramakar (2014) A study on selection of probability distributions for at-site flood frequency analysis in Mahanadi River Basin, India. Proceedings of the International Conference on Fluvial Hydraulics, River Flow, pp. 1813-1819.

Guru, N. and Jha, R. (2016) Flood estimation in Mahanadi river system, India using partial duration series. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, v. 10(2), pp. 135-145.

Haddad, K. and Rahman, A. (2011) 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, v. 25(3), pp. 415-428.

Karim, M. A. and Chowdhury, J. U. (1995) A comparison of four distributions used in flood frequency analysis in Bangladesh. Hydrological Sciences Journal, v. 40(1), pp. 55-66.

Khatua, K. K. and Patra, K. C. (2004) Management of high flood in Mahanadi and Its tributaries below Naraj. 49th Annual session, IEI, Orissa State Centre, Bhubaneswar.

Krishna, G. S. and Veerendra, G. (2015) Flood frequency analysis of Prakasam barrage reservoir Krishna District, Andhra Pradesh using Weibull, Gringorten and L-moments formula. International Journal of Civil, Structural, Environmental and Infrastructure Engineering Research and Development, v. 5(2), pp. 57-62.

Kumaraswamy, P. (1980) A generalized probability density function for double-bounded random processes. Journal of Hydrology, v. 46(1-2), pp. 79-88.

Langat, P. K., Kumar, L. and Koech, R. (2019) Identification of the most suitable probability distribution models for maximum, minimum, and mean streamflow. Water, v. 11(4), pp. 734.

Nagesh, S. and Laxmi, Dharmannavar (2021) Identifying Suitable Probability Models for Extreme Flood Heights in Mahanadi River Basin. International Journal of Agricultural and Statistical Science, v. 17(1), pp. 197-207.

Nayak, S. and Mandal, M. (2019) Assessment of Sea Level and Morphological Changes along Indian Coastal Areas during 1975-2005. Earth Science India, v. 12(2), pp. 117-125.

Naz, S., Baig, M. J., Inayatullah, S., Siddiqi, T. A., & Ahsanuddin, M. (2019) Flood Risk Assessment of Guddu Barrage using Gumbel's Distribution. International Journal of Sciences, v. 8(04), pp. 33-38.

Pai-Panandiker, A., Gude, S., Chachadi, A. G. and Mahender, K. (2020) Evaluating the effect of land-use change coupled with climate change: A study based on Sal river flow in Goa, India. Earth Science India, v. 13 (4), pp. 96-106.

Sharma, A., Nikam, B. R., Bagri, D. S. and Kumar, D. (2020) Trend Analysis of Annual and Seasonal Time Series over the Last 46 Years in Asan Watershed, Doon Valley Based on Gridded Data Set. Earth Science India, v. 13(1), pp. 8-19.

Singo, L. R., Kundu, P. M., Odiyo, J. O., Mathivha, F. I., and Nkuna, T. R. (2012) Flood frequency analysis of annual maximum stream flows for Luvuvhu River catchment, Limpopo Province, South Africa. 16th SANCIAHS National Hydrology Symposium, 1-3 October 2012, University of Pretoria, Pretoria.

Smithers, J. C. (2012) Methods for design flood estimation in South Africa. Water SA, 38(4), 633-646.

Solomon, O. and Prince, O. (2013) Flood frequency analysis of Osse river using Gumbel's distribution. Civil and Environmental Research, v. 3(10), pp. 55-59.

Sukla, M. K., Mungaraj, A. K. and Sahoo, L. N. (2014) An investigation on the stochastic modelling of daily rainfall amount in the Mahanadi Delta region, India. Research Journal of Mathematical and Statistical Sciences, v. 2 (9), pp. 1–8.

Sundar, V. and Subbiah, K. (1989) Application of double bounded probability density function for analysis of ocean waves. Ocean Engineering, v. 16(2), pp. 193-200.

Swetapadma, S. and Ojha, C. S. P. (2017) Sustainable River Basin Management Planning: A Holistic Approach. Earth Science India, v. 10, pp. 126-152.

Vivekanandan, N. (2015) Frequency analysis of annual maximum flood discharge using method of moments and maximum likelihood method of gamma and extreme value family of probability distributions. International Journal of Mathematics and Computational Science, v. 1(3), pp. 141-146.

(Received: 07/07/2021; Accepted: 28.08.2021)