

# Optimal Probability Distribution Function and Plotting Position for Annual Maximum Flood in Mahanadi River System, India

Rajesh Ranjan <sup>1</sup>, Ramakar Jha <sup>2\*</sup>

<sup>1</sup> Department of Civil Engineering, National Institute of Technology, Patna, India. E-mail: [rajeshr.phd19.ce@nitp.ac.in](mailto:rajeshr.phd19.ce@nitp.ac.in)

<sup>2</sup> Department of Civil Engineering, National Institute of Technology, Patna, India. E-mail: [rj@nitp.ac.in](mailto:rj@nitp.ac.in)

ARTICLE INFO	ABSTRACT
Received: 29 Dec 2024	<p>In the present work, optimum probability distribution function and plotting position of annual maximum flood series data obtained for Mahanadi River system has been analysed. For the analysis, 25 to 45 years of data obtained from Nineteen gauging stations of Mahanadi river system have been considered. Different commonly used probability distribution functions have been tested for their applicability at all the nineteen gauging stations individually. They are Generalised Extreme Value (GEV), Generalised Pareto (GP), Log-Pearson III (LP III), Normal, Log normal (LN), Log-Logistic (LL), Log-Logistic 3P(LL-3P), Log normal 3P(LN-3P), Pareto (P) and Gumbel max method. A comparison of the results obtained using these distributions has been done with the Weibull, Hazen &amp; California plotting positions. From the results, the Generalised Pareto (GP) probability distribution provided best results followed by GEV and Log Pearson Type III. The goodness of fit tests namely Kolmogorov-Smirnov (KS), Anderson-Darling (AD), and Chi squared (CS) tests has been done to assess the best performing probability distribution. The annual maximum flow observed are compared well with the predicted annual maximum flow for the return period of 5, 10, 20, 25, 30, 35 and 45 years. The results obtained are promising and can be used for design and development of water resources projects in Mahanadi river system.</p> <p><b>Keywords:</b> Annual Maximum Flood (AMF), Flood Frequency Analysis, Hydrological Modeling, Flood Probability Distribution, Flood Discharge Distribution, Mahanadi River Floods, Gumbel Distribution for Floods.</p>
Revised: 12 Feb 2025	
Accepted: 27 Feb 2025	

## INTRODUCTION

Natural catastrophes like floods can result in fatalities, serious injuries, devastation of agricultural areas, and significant property losses (Fill and Stedinger, 1995). Implementing flood frequency analysis to identify efficient hydraulic structure designs is one way to reduce flood damages and monetary losses. Estimating peak discharges for design purposes on catchments with inadequate information available has been a persistent issue in hydrology (Blazkova and Beven, 1997). An effective and promising solution to this issue is the calculated flood frequency curve. However, historical data required to estimate these statistics are not always available at the site of interest, or the data that is available may not be representative of the basin flow due to changes in the watershed characteristics, such as urbanization. Reliable estimates of flow statistics, such as mean annual flow and flood quantiles, are required (Pandey and Nguyen, 1999; Ouada, et al, 2006).

In actuality, single-site and/or regional flood-frequency studies are frequently used to estimate design floods (Burn, 1990). Appropriate flood frequency and risk analyses can lead to an ideal design (Saf, 2008). However, modelling and sampling errors might affect design floods computed using fitted distributions (Alila and Mtiraoui, 2002). Many researchers have studied various kinds of distributions for use in flood-frequency analysis (Cunnane, 1989; GREHYS, 1996; Blazkova and Beven, 1997; Saf, 2008).

In particular in developing nations, the historical hydrometric data that is now available may be sparse, non-existent, or so limited that it is not typical of the region in issue, or obtaining it may be costly, challenging, or time-consuming (Öztekin, Karaman and Brown, 2007; Patel, 2007). The most common application of statistics in hydrology worldwide has been frequency analysis, primarily used for estimating flood flows. The most suitable probability distributions for different situations are determined by the specific properties and characteristics of those distributions (Haan, 1994). It is challenging for hydrologists to accurately predict flood levels based on the limited historical data on runoff, rainfall, and river stages.

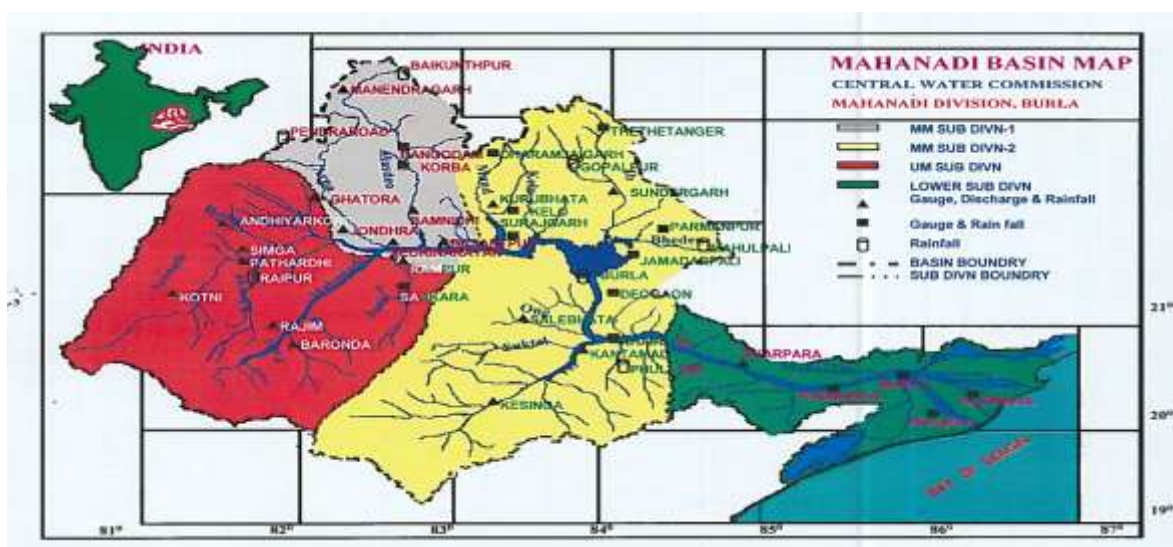
To fit flood extreme data, a variety of parameter (Singh and Strupczewski, 2002). Öztekin, Karaman and Brown (2007) estimate techniques have been employed to a wide range of distributions. Around the world, various research on distribution selection for flood data have been conducted. While the generalised extreme value distribution is used in Great Britain and the lognormal distribution is used in China, the three-parameter log-Pearson type 3 distribution is the most commonly utilised distribution in the United States (Singh and Strupczewski, 2002). Various distributions of floods have also been researched, for example in USA (Wallis, 1988; Vogel, Thomas and McMahon, 1993); UK, Australia, Italy Scotland, Turkey and Kenya (Haktanir, 1991; Mutua, 1994; Abdul Karim and Chowdhury, 1995). It is undeniable that different places may have distinct hydro-climatic regimes; yet, these variations ought to provide a hydro-physical foundation for selecting a specific distribution. As a result, choosing a suitable distribution requires further consideration.

In hydrology, probability distribution functions of continuous random variables are commonly used to fit data distributions. While all plotting position methods yield similar values near the centre of the distribution, significant differences can occur in the tails (Hann, 1994). Chow (1964) presents a number of plotting position correlations while Haan (1994) recommended using the three often used relationship positions that meet the Gumbel (1958) five requirements for charting position relationships: California, Hazen, and Weibull. Similarly, Abida and Ellouze (2007) opined that the most commonly applied distributions now being the Gumbel (EV1), the Generalized Extreme Value (GEV), the Log Pearson Type III (LP3), and the Three parameter Lognormal (LN3). A detailed FFA for Mahanadi River basin has been done for 19 gauging stations and 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-Smirnov (KS), Anderson-Darling (AD), and Chi squared (CS) tests (R Ranjan & Ramakar Jha 2024)

Three popular distributions (mainly Generalised Extreme value distribution, Generalised Pareto & Log Pearson III) and three distinct plotting position relationships were utilised in this study to determine which flood frequency distribution best matches in Mahanadi river basin for all nineteen gauge stations.

## **THE STUDY AREA AND DATA COLLECTION**

The Mahanadi is one of the major peninsular rivers of the country and, in water potential and flood producing capacity; it ranks after Godavari. The basin is broadly divided into three sub-basins; Mahanadi Upper Sub Basin, Mahanadi Middle Sub Basin and Mahanadi Lower sub basin. The Mahanadi basin extends over five states of Chhattisgarh and Odisha and comparatively smaller portions of Jharkhand, Maharashtra and Madhya Pradesh, draining an area of 141,589 Sq.km. It is the 8<sup>th</sup> largest basin and a major river of east central India, which is nearly 4.28% of the total geographical area of the country (Figure 1). The geographical extent of the basin lies between 80°28' and 86°43' east longitudes and 19°8' and 23°32' north latitudes.



**Figure 1.** The study area of Mahanadi river basin (Source: CWC Water Year Book)

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 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.

## METHODOLOGY

Table 1 provides the probability distribution functions of three different methods and the governing equations use for the estimation of optimum distribution function for annual maximum flood series.

**Table 1:** Governing equations used in three different methods provided best results

Distributions	Probability Density Function	Cumulative Distribution Function
Generalized Extreme Value	$f(x) = \begin{cases} \frac{1}{\sigma} \exp\left(-(1+kz)^{\frac{-1}{k}}\right) (1+kz)^{-1-\frac{1}{k}} & ; k \neq 0 \\ \frac{1}{\sigma} \exp(-z - \exp(-z)) & ; k = 0 \end{cases} \quad (8)$	$F(x) = \begin{cases} \exp\left(-(1+kz)^{\frac{-1}{k}}\right) & ; k \neq 0 \\ \exp(-\exp(-z)) & ; k = 0 \end{cases} \quad (9)$
Generalised Pareto	$f(x) = \begin{cases} \frac{1}{\sigma} \left(1 + k \frac{x-\mu}{\sigma}\right)^{\frac{1}{k}-1} & ; k \neq 0 \\ \frac{1}{\sigma} \exp\left(-\frac{x-\mu}{\sigma}\right) & ; k = 0 \end{cases} \quad (12)$	$F(x) = \begin{cases} 1 - \left(1 + k \frac{x-\mu}{\sigma}\right)^{-\frac{1}{k}} & ; k \neq 0 \\ 1 - \exp\left(-\frac{x-\mu}{\sigma}\right) & ; k = 0 \end{cases} \quad (13)$
Log Pearson Type III	$f(x) = \frac{1}{x \beta \Gamma(\alpha)} \left(\frac{\ln x - \gamma}{\beta}\right)^{\alpha-1} \exp\left(-\frac{\ln x - \gamma}{\beta}\right) \quad (14)$	$F(x) = \frac{\Gamma(\ln x - \gamma)(\alpha)}{\beta \Gamma(\alpha)} \quad (15)$

(k, α) are shape, (σ,β) are scale and (μ,γ) are location parameter;  $\varphi$ =Laplace Integral;  $\Gamma$ =gamma function

The annual maximum flood series data have been used to establish the plotting position of the magnitude of the data and likelihood of exceeding it. Table 2 provides the methods used for the analysis of plotting position relationships.

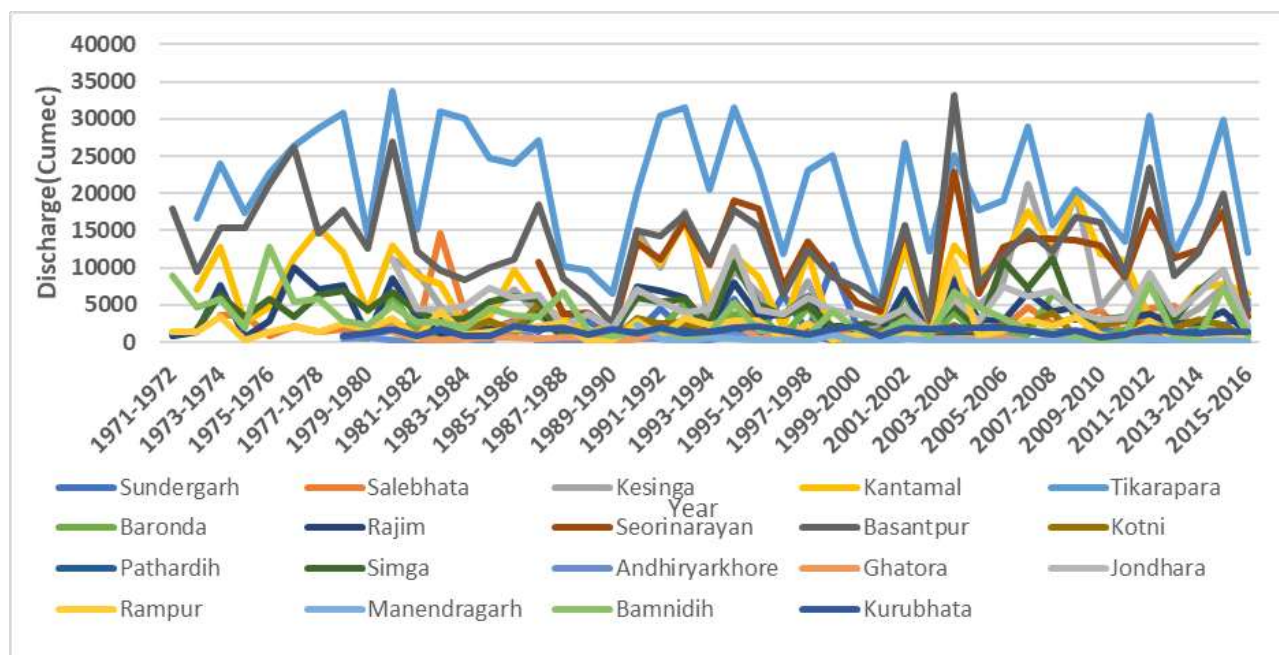
**Table 2:** Plotting Position relationships

Name	Source	Relationship
California	California (1923)	$\frac{m}{n}$
Hazen	Hazen (1930)	$\frac{2m - 1}{2n}$
Weibull	Weibull (1939)	$\frac{m}{n + 1}$

Where m= Rank; N= number of years of record.

## RESULTS AND DISCUSSIONS

The time series plot of maximum or peak annual flows of 19 stations for period of 26 years to 45 years is shown in Figure 2. The highest discharge is observed to be 33800 m<sup>3</sup>/s in 1980-81 for Tikarapara and lowest discharge is 85 m<sup>3</sup>/s in 2013-14 for Manendragarh.

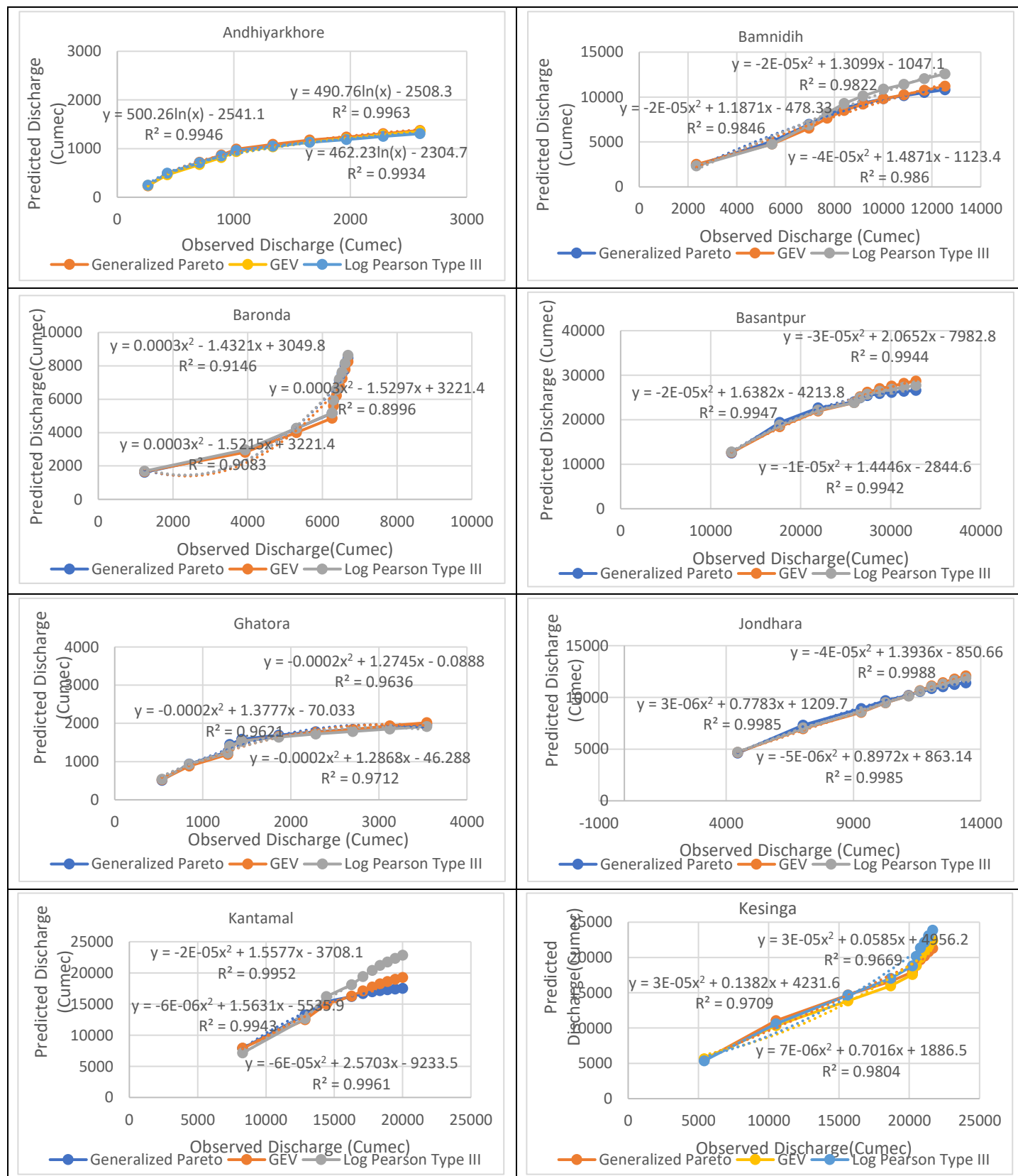


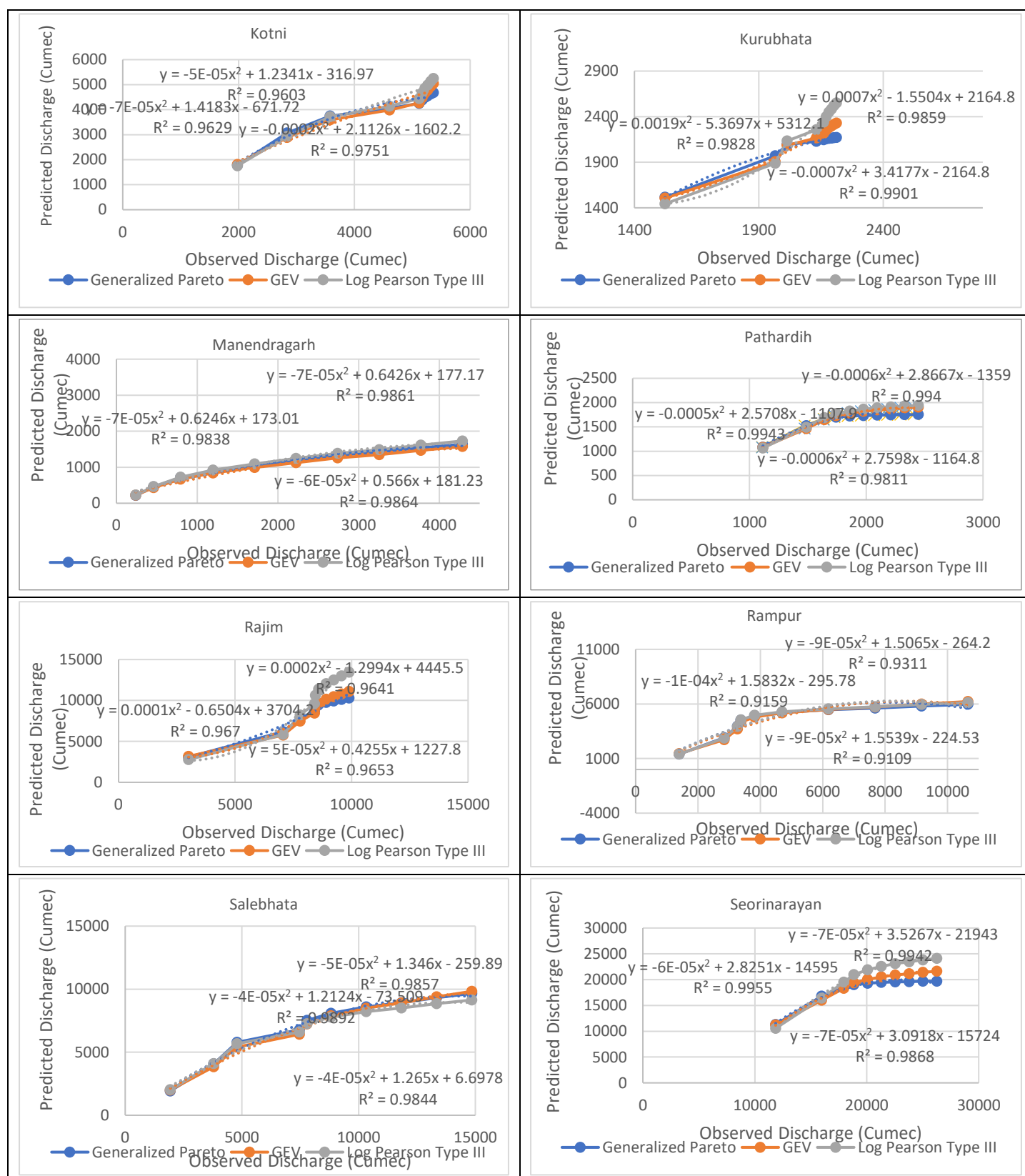
**Figure 2.** Annual maximum discharge of all the stations lying in the study area

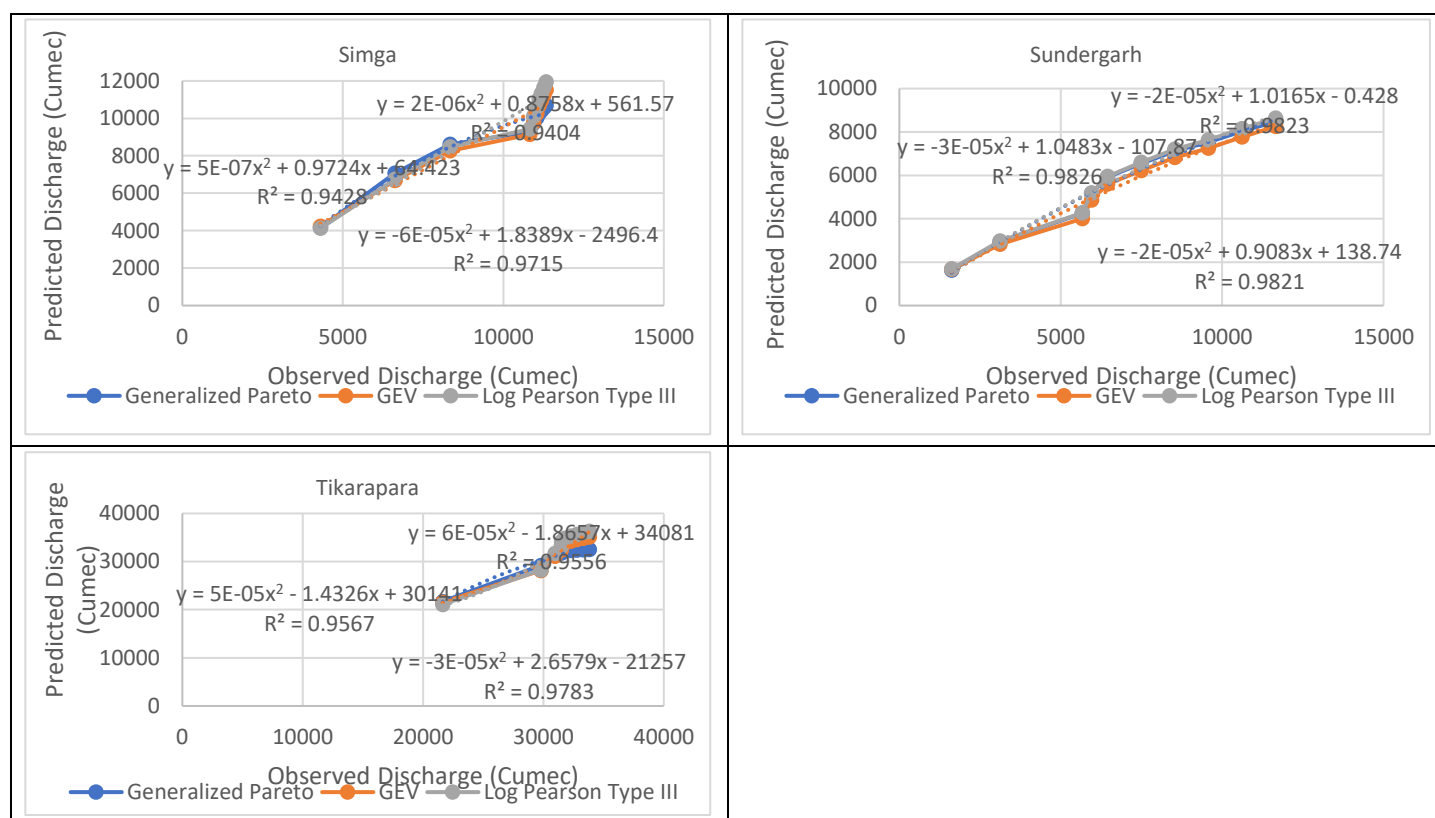
All the probability distribution functions have been used to analyse the best distribution applicable for all the nineteen discharge gauging stations. A comparison of different distributions have been compared with the Weibull's plotting position, Hazen's plotting position and California's plotting position. Three probability distributions namely Generalised Extreme Value (GEV), Generalized Pareto (GP) and Log Pearson Type III (LP-III) provided best results as per different goodness of fit tests.

In some cases, the results obtained from probability distribution functions GEV, GP and LP-III are found to underestimate or overestimate the annual flood series. The results for Weibull's, Hazen's and California's plotting positions are shown in Figure 3, Figure 4 and Figure 5 respectively.

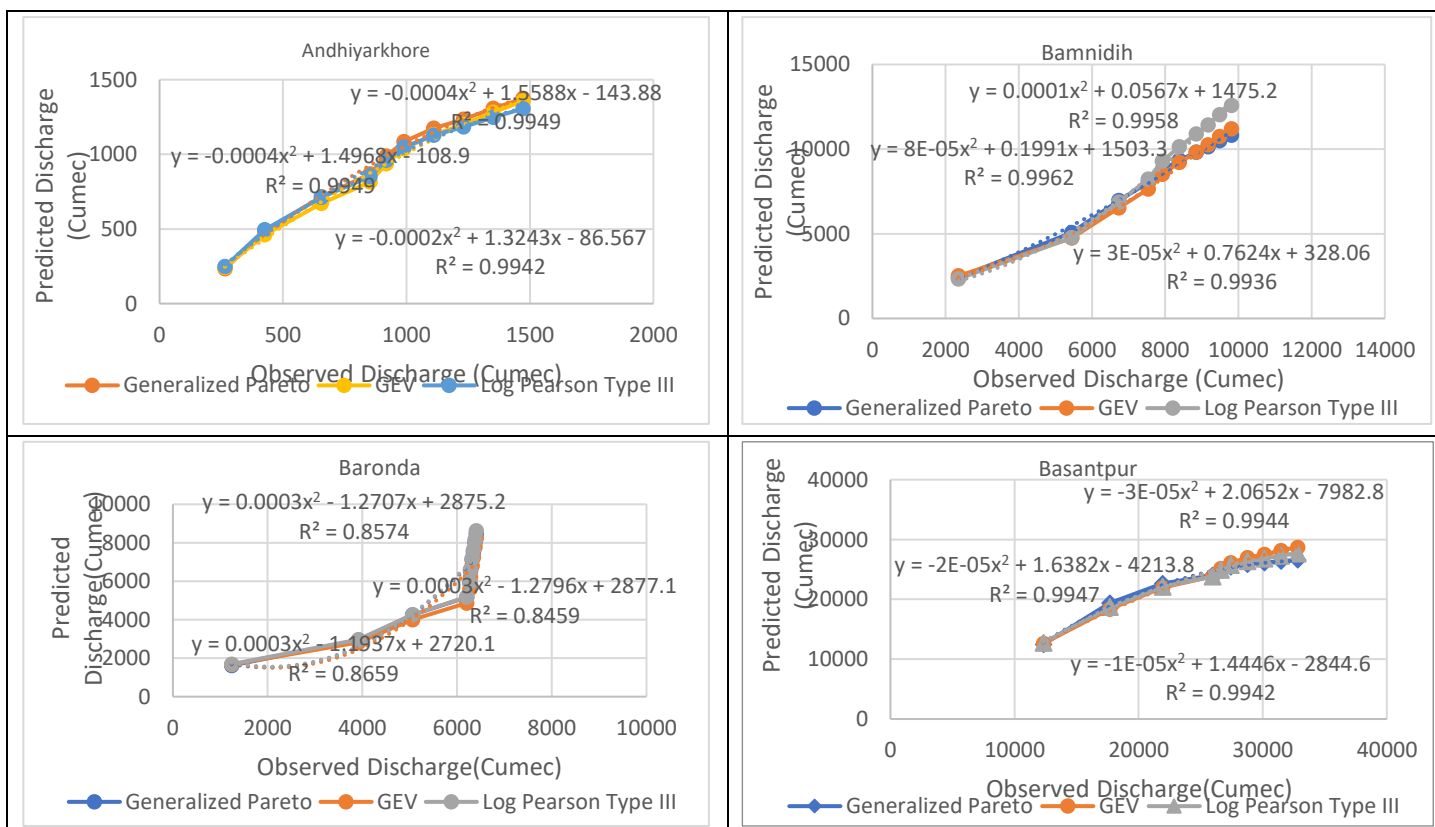


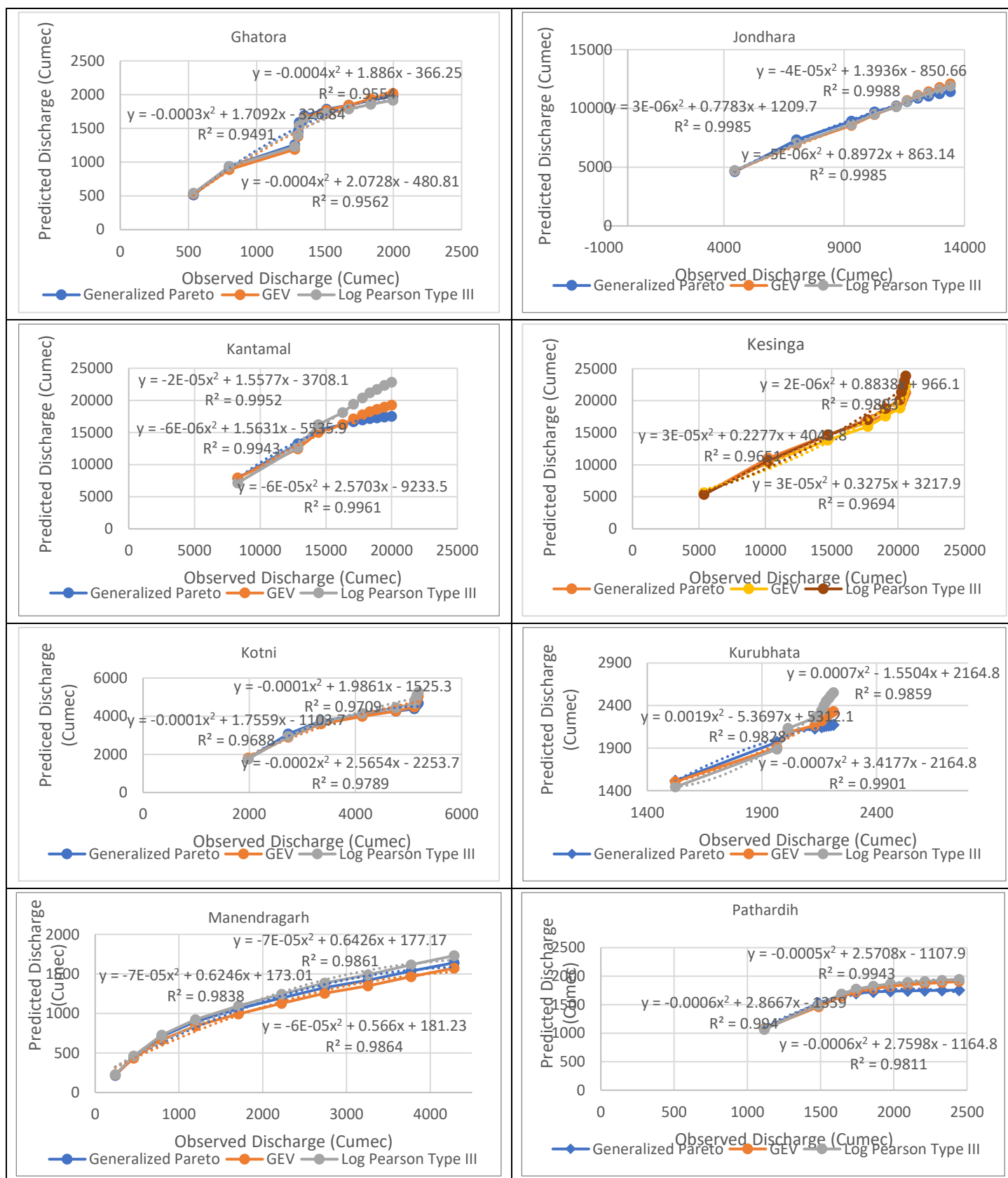




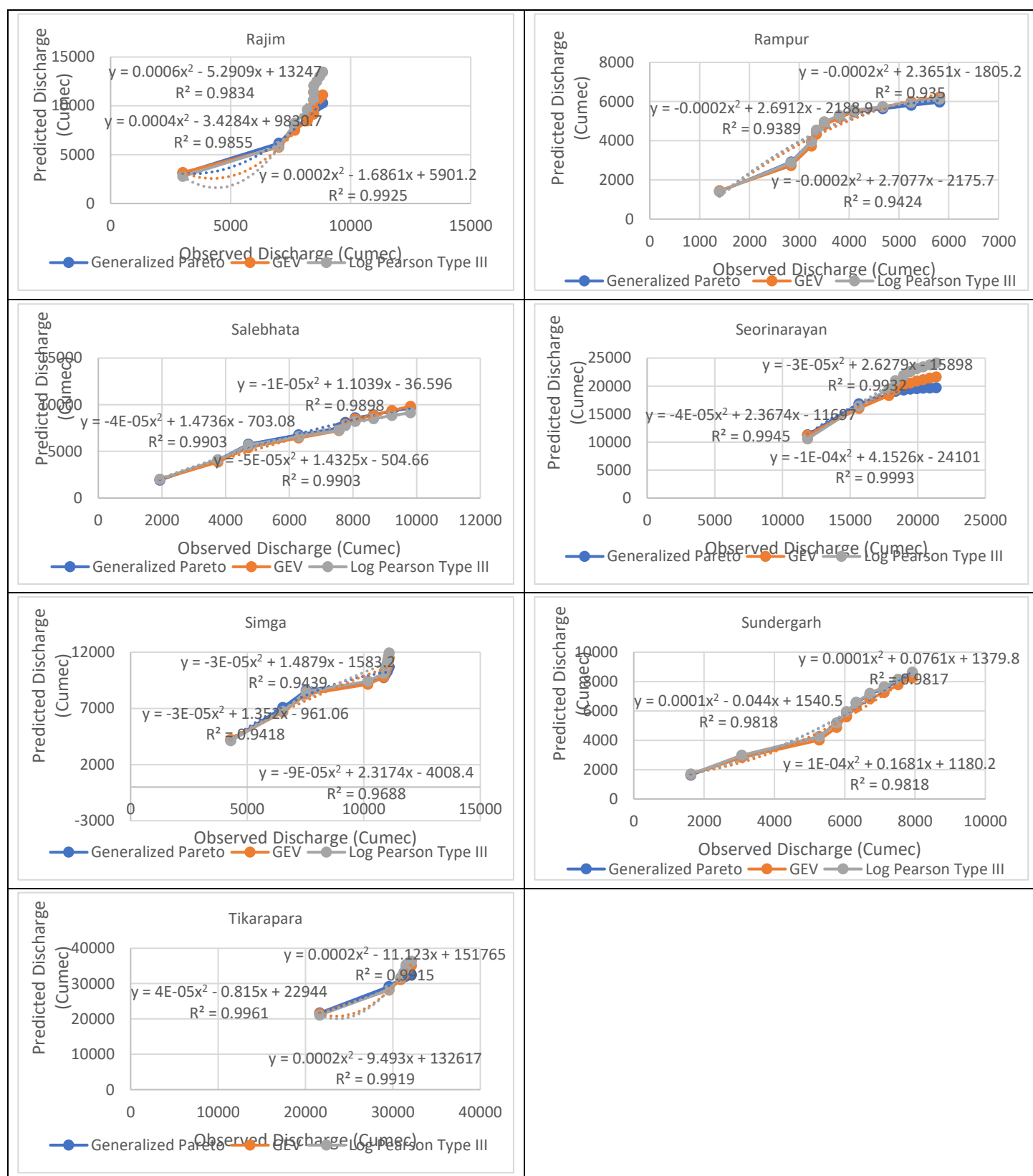


**Figure 3.** A comparison of probability distribution fuctions with Weibull's plotting position

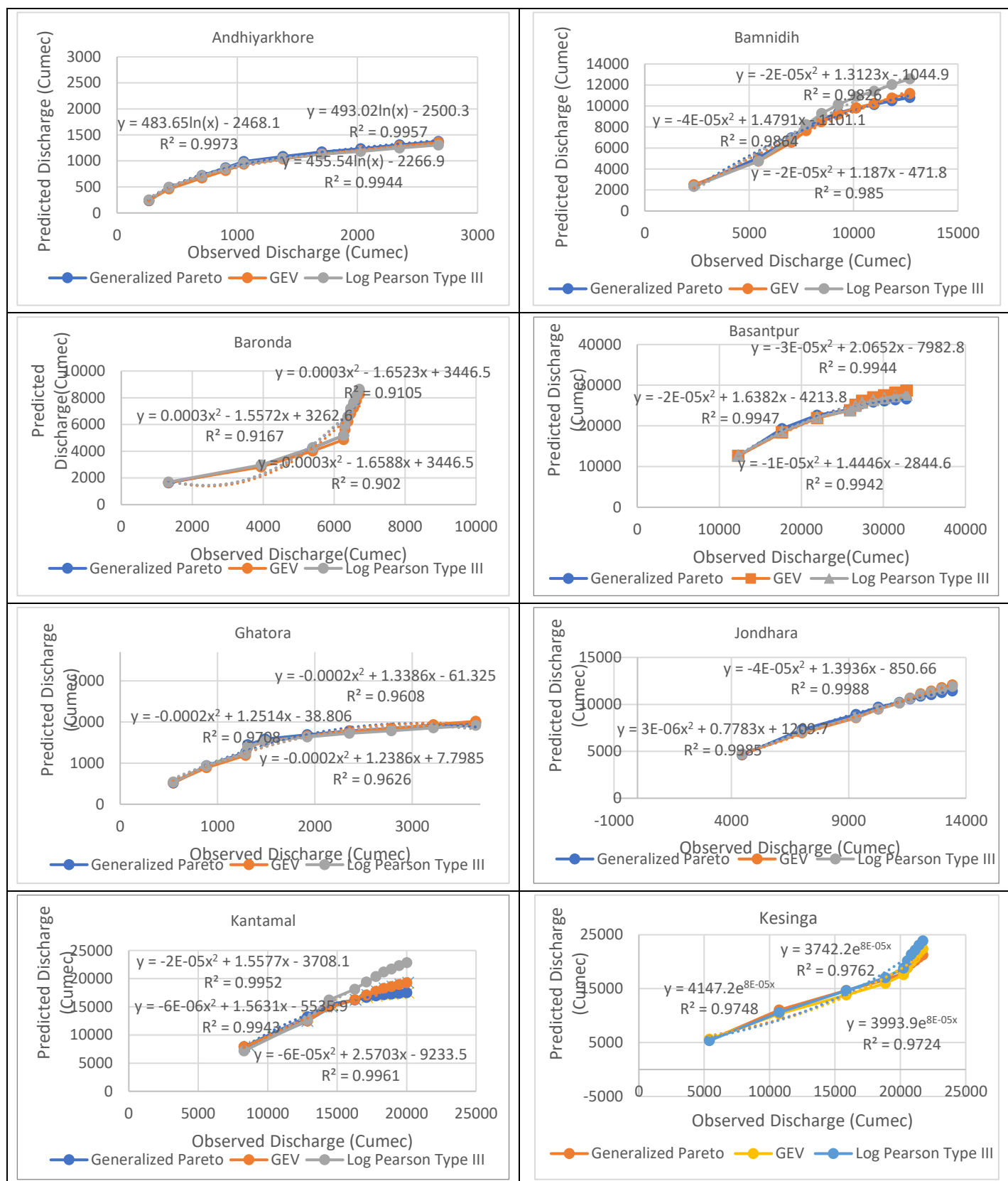


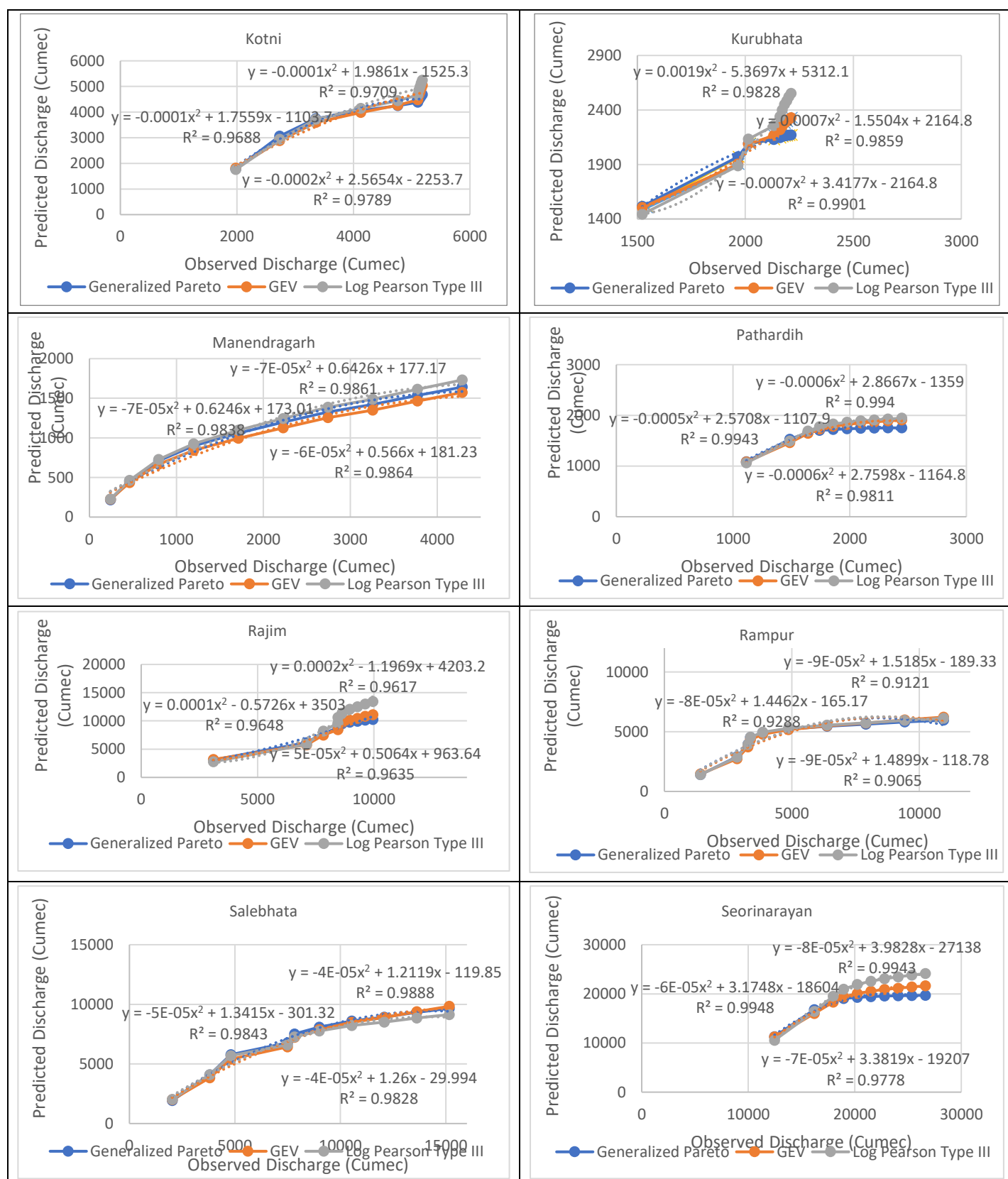


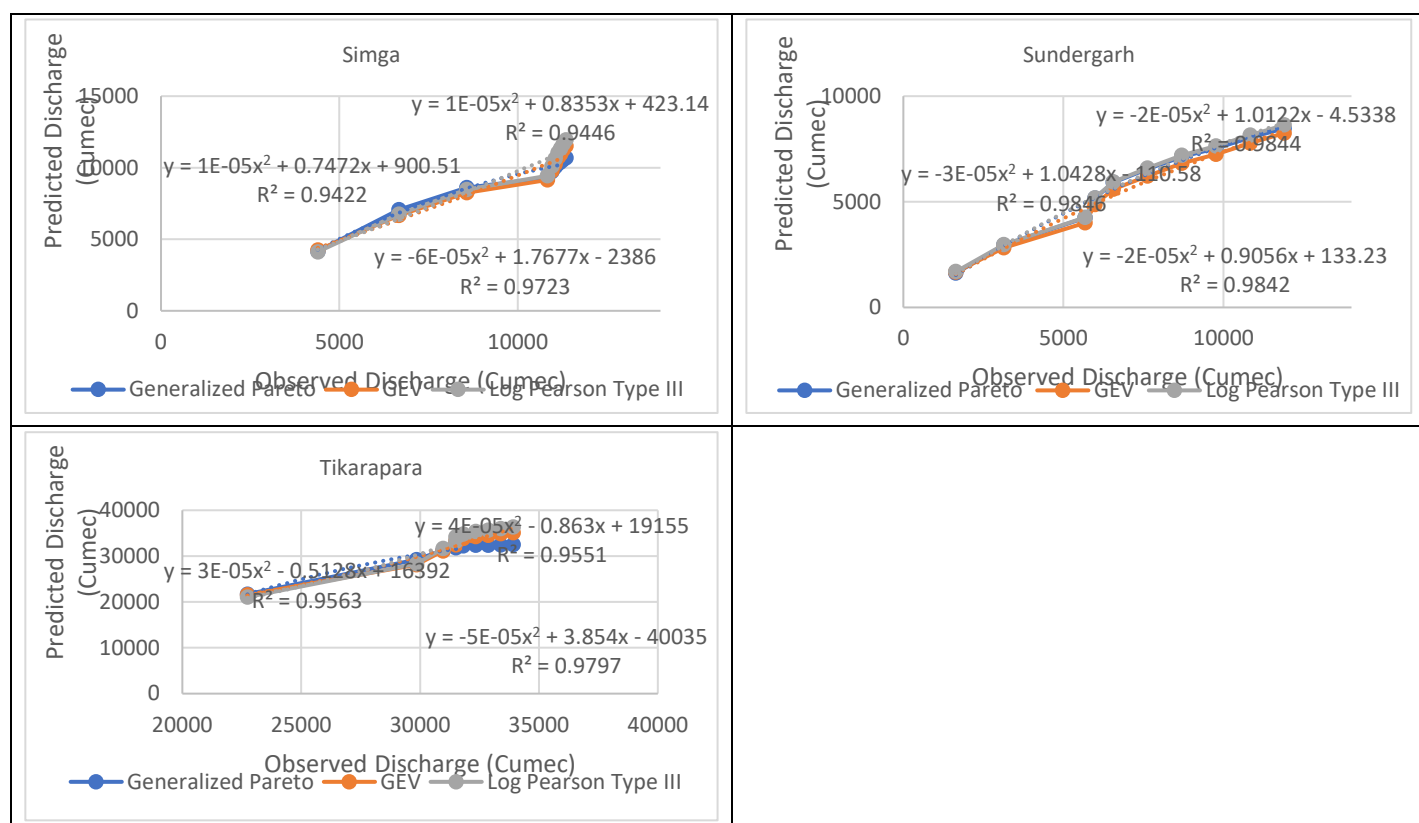




**Figure 4.** A comparison of probability distribution fucntions with Hazen's plotting position







**Figure 5.** A comparison of probability distribution functions with California's plotting position

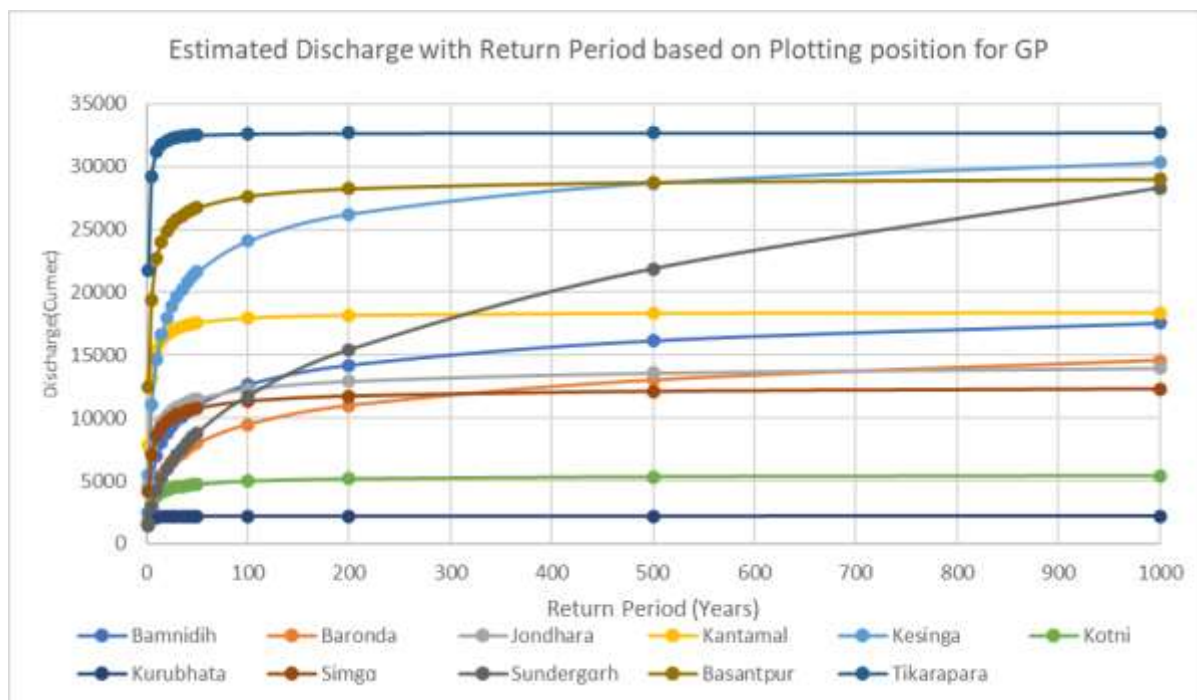
The coefficient of determination ( $R^2$ ) is a widely used statistical metric in regression analysis, representing the proportion of variance in the dependent variable that is explained by the independent variable(s) (Draper & Smith, 1998). It serves as an indicator of model performance, with values ranging between 0 and 1, where higher values suggest a better fit of the model to the observed data. Table 3 shows the coefficient of determination ( $R^2$ ) for the 19 gauging stations for Mahanadi river Basin for the three different plotting position (Weibull, Hazen, California) for three best distribution outcome of goodness of fit i.e GEV, Generalised Pareto & log Pearson III. It has been seen that GP is best for the 10 gauging station (Bamnidihi, Baronda, Jondhara, Kantamal, Kesianga, Kotni, Kurubhata, Simga, Sundergarh & Tikarapara) having maximum coefficient of determination ( $R^2$ ) Whereas Generalised Extreme Value distribution is best for 8 station (Andhiyarkhore, Ghatara, Manendragarh, Pathardihi, Rajim, Rampur, Salebhata & Seorinarayan) and Log Pearson III is best for Basantpur considering maximum coefficient of determination ( $R^2$ ) for minimum two or three plotting position.

**Table 3:** Coefficient of Determination( $R^2$ ) for different gauging stations using three (Weibull, Hazen, California) Plotting Position for Mahanadi River Basin.

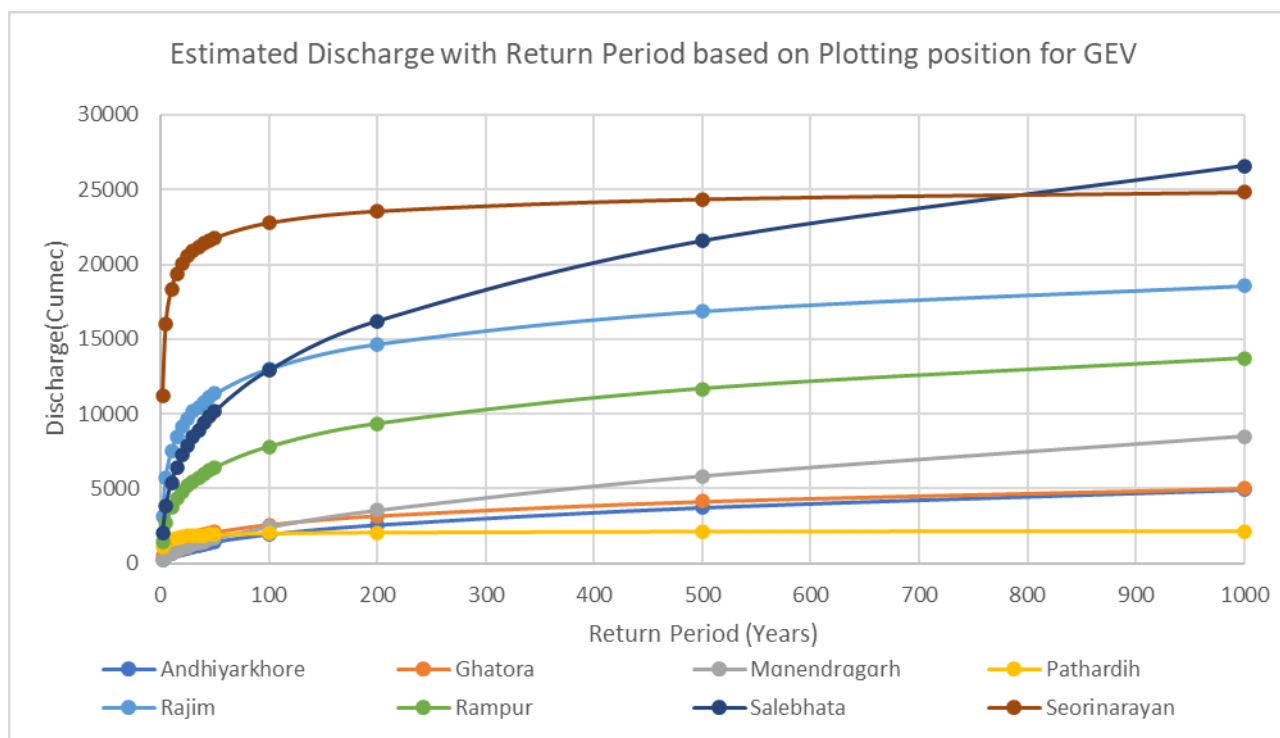
SI NO.	Site Name	Weibull			Hazen			California		
		Generaliz ed Pareto	GEV	Log Pearson Type III	Generaliz ed Pareto	GEV	Log Pearson Type III	Generaliz ed Pareto	GEV	Log Pearson Type III
		$R^2$	$R^2$	$R^2$	$R^2$	$R^2$	$R^2$	$R^2$	$R^2$	$R^2$
1	Andhiyarkore	0.9946	0.9963	0.9934	0.9949	0.9942	0.9949	0.9957	0.9973	0.9944
2	Bamnidihi	0.986	0.9846	0.9822	0.9936	0.9962	0.9958	0.9864	0.985	0.9826
3	Baronda	0.9146	0.8996	0.9083	0.8659	0.8459	0.8574	0.9167	0.902	0.9105
4	Basantpur	0.9944	0.9944	0.9947	0.9935	0.9877	0.99	0.9935	0.9934	0.994
5	Ghatora	0.9621	0.9712	0.9636	0.9562	0.9491	0.9554	0.9608	0.9708	0.9626
6	Jondhara	0.9988	0.9985	0.9985	0.9975	0.9942	0.995	0.999	0.9986	0.9987
7	Kantamal	0.9961	0.9952	0.9943	0.9896	0.9884	0.9878	0.9979	0.9972	0.9962
8	Kesinga	0.9804	0.9669	0.9709	0.9803	0.9651	0.9694	0.9748	0.9724	0.9762
9	Kotni	0.9751	0.9603	0.9629	0.9789	0.9688	0.9709	0.9725	0.958	0.9606
10	Kurubhata	0.9901	0.9859	0.9828	0.9897	0.9826	0.976	0.9915	0.9882	0.9855
11	Manendragarh	0.9835	0.9861	0.9857	0.9994	0.9996	0.9996	0.9838	0.9864	0.9861
12	Pathardih	0.9831	0.995	0.9949	0.9992	0.993	0.9935	0.9811	0.9943	0.994
13	Rajim	0.9653	0.967	0.9641	0.9925	0.9855	0.9834	0.9635	0.9648	0.9617
14	Rampur	0.9109	0.9311	0.9159	0.9424	0.935	0.9389	0.9065	0.9288	0.9121
15	Salebhata	0.9857	0.9892	0.9844	0.9903	0.9898	0.9903	0.9843	0.9888	0.9828
16	Seorinaryan	0.9868	0.9955	0.9942	0.9993	0.9945	0.9932	0.9778	0.9948	0.9943
17	Simga	0.9715	0.9404	0.9428	0.9688	0.9418	0.9439	0.9723	0.9422	0.9446
18	Sundergarh	0.9826	0.9821	0.9823	0.9818	0.9818	0.9817	0.9846	0.9842	0.9844
19	Tikarapara	0.9783	0.9567	0.9556	0.9961	0.9919	0.9915	0.9797	0.9563	0.9551

Based on the best Plotting position GP is best for 11 locations(Bamnidihi, Baronda, Jondhara, Kantamal, Kesinga, Kotni, Kurubhata, Simga, Sundergarh, Basantpur & Tikarapara). Graph(Fig. 6) showing the estimated discharge obtained from 2, 5, 10, 25, 30, 35, 40 & 45 years return period predicted for 50, 100, 200, 500 & 1000 years is shown flow is taken into account. GEV is best for 8 locations(Andhiyarkore, Ghatora, Manendragarh, Pathardih, Rajim, Rampur, Salebhata & Seorinarayan) Graph(Fig.7) is shown for different return period predicted for the same below.





**Figure 6.** Predicted discharge for GP for best plotting position for 11 stations



**Figure 7.** Predicted discharge for GEV for best plotting position for 8 stations

## CONCLUSIONS

The following conclusions are made:

- The annual maximum flow data (26-45 years) used for the analysis have been tested for its applicability by using 10 different probability distribution function. Three probability distribution functions namely

Generalised Extreme Value (GEV), Generalized Pareto (GP) and Log Pearson Type III (LP-III) are found to provide best results.

- The Plotting position methods have been used to compare the results obtained using GEV, GP and LP-III. These methods are Weibull's, Hazen's and California's plotting positions.
- The coefficient of determination ( $r^2$ ) is found to be very high for GEV, GP and LP-III distribution functions for all the 19 discharge gauging stations.
- The minimum absolute differences at return periods of 5, 10, 25 and 45 years were obtained under the GEV distribution when matched with Hazen plotting position.

The probability distribution function are very useful tools for design of water resources structures.

### REFERENCES

- [1] Abdul-Karim, M. D., and J. U. Chowdhury. 1995. A comparison of four distributions used in flood frequency analysis in Bangladesh. *Hydrol. Sci. J.*, 40(1): 55-66
- [2] Abida, H., and M. Ellouze. 2007. Probability distribution of flood flows in Tunisia. *Hydrol. Earth Syst. Sci. Discuss.*, 4: 957–981. [www.hydrol-earth-syst-sci-discuss.net/4/957/2007/](http://www.hydrol-earth-syst-sci-discuss.net/4/957/2007/)
- [3] Adeboye, O. B., and M. O. Alatise. 2007. Performance of probability distributions and plotting positions in estimating the flood of river Osun at Apoje Sub-basin, Nigeria. *Agricultural Engineering International: the CIGR Ejournal*. Manuscript LW 07 007. Vol. IX.
- [4] Alila, Y., and A. Mtiraoui. 2002. Implications of heterogeneous flood-frequency distributions on traditional stream-discharge prediction techniques. *Hydrol. Process*, 16: 1065–1084. DOI: 10.1002/hyp.346
- [5] Blazkova, S., and K. Beven. 1997. Flood frequency prediction for data limited catchments in the Czech Republic using a stochastic rainfall model and TOPMODEL. *Journal of Hydrology*, 195: 256-278.
- [6] Burn, D. H. 1990. Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research*, 26(10): 2257–2265.
- [7] Chin, D. A. 2006. *Water-Resources Engineering*. 2nd Edition. Pearson Prentice Hall, Upper Saddle River, New Jersey.
- [8] Chow, V. T. 1964. *Handbook of Applied Hydrology*. McGraw-Hill, New York.
- [9] Cunnane, C. 1989. Statistical distributions for flood frequency analysis. WMO Operational Hydrology Report no. 33. World Meteorological Organisation, 41 Giuseppe Motta, Casa Postale no. 5, CH 1211 Geneva, Switzerland.
- [10] Eze, B. E. 1997. Rain splash detachment on different land use surfaces in a humid tropical environment: a case study of Ibadan. Unpublished PhD Thesis, Department of Geography, University of Ibadan, Ibadan, Nigeria.
- [11] Fill, H. D., and J. R. Stedinger. 1995. Homogeneity tests based upon Gumbel distribution and a critical appraisal of Dalrymple's test. *Journal of Hydrology*, 166: 81-105
- [12] GREHYS (Groupe de recherche en hydrologie statistique). 1996. Inter-comparison of regional flood frequency procedures for Canadian rivers. *Journal of Hydrology*, 186: 85-103.
- [13] GREHYS (Groupe de Recherche en Hydrologie Statistique). 1996. Presentation and Review of Some Methods for Regional Flood Frequency Analysis. *Journal of Hydrology*, 186: 63-84.
- [14] Gumbel, E. J. 1958. *Statistics of extreme values*. Columbia University Press. New York.
- [15] Haan, C. T. 1994. *Statistical methods in hydrology*. Iowa State University Press. Ames.
- [16] Haktanir, T. 1991. Statistical modelling of maximum flows in Turkish rivers. *Hydrol. Sci. J.*, 36(4): 367-389.
- [17] Joshua, E. O., and O. A. Oyebanjo. 2009. Distribution of heavy metals in sediments of Osun river basin southwestern Nigeria. *Research Journal of Earth Sciences*, 1(2): 74-80.
- [18] Kumar N & Jha R 2023, Flood Frequency Analysis for Kosi River Basin, Bihar, India Using Statistical Methods, *Civil Engineering and Architecture* 12(2): 991-999, 2024
- [19] Lal, R. 1993. Tillage effects on soil degradation, soil resilience, soil quality, and sustainability. *Soil Tillage Research*, 27: 1 - 8.
- [20] Mutua, F. M. 1994. The use of the Akaike information criterion in the identification of an optimum flood frequency model. *Hydrol. Sci. J.*, 39(3): 235-244.

- [21] Ouarda, T. B., M. J. Cunderlik, A. St-Hilaire, M. Barbet, P. Bruneau, and B. Bobe'e. 2006. Data-based comparison of seasonality-based regional flood frequency methods. *Journal of Hydrology*, 330: 329– 339.
- [22] Öztekin, T., S. Karaman, and L. C. Brown. 2007. The best distributions for maximum and minimum daily temperatures of Columbus, Ohio. ASABE Paper No. 072298. St. Joseph, Mich.: ASABE.
- [23] Pandey, G. R., and V. T. V. Nguyen. 1999. A comparative study of regression based methods in regional flood frequency analysis. *Journal of Hydrology*, 225: 92–101.
- [24] Patel, J. A. 2007. Evaluation of Low Flow Estimation Techniques for Ungauged Catchments. *Water and Environment Journal*, 21: 41–46.
- [25] Ranjan R & Jha R 2024, Flood Frequency Analysis for Water Resources Planning and Management in Mahanadi River System, India, *Nanotechnology Perceptions* 20 No. S16 (2024) 1474-1490.
- [26] Saf, B. 2008. Application of index procedures to flood frequency analysis in Turkey. *Journal of the American Water Resources Association (JAWRA)*, 44(1): 37-47. DOI: 10.1111/j.1752-1688.2007. 00136.x
- [27] Singh, V. P., and W. G. Strupczewski. 2002. On the status of flood frequency analysis. *Hydrol. Process.*, 16: 3737–3740. DOI: 10.1002/hyp.5083
- [28] Vogel, R. M., W. O. Thomas, and T. A. McMahon. 1993. Flood-flow frequency model selection in the south western United States. *Wat. Resour. Plan. Manage. ASCE*, 119(3): 353-366.
- [29] Wallis, J. R. 1988. Catastrophes, computing and containment: living in our restless habitat. *Speculation in Science and Technology*, 11(4): 295-315.