Application of Gumbel's Distribution Method for Flood Frequency Analysis of Lower Ganga Basin (Farakka Barrage Station), West Bengal, India

Mandal Kajal Kumar[1*], Dharanirajan K.[1] and Sarkar Sabyasachi[2]
1. Department of Disaster Management, Pondicherry University, Port Blair Campus, A and N Islands, INDIA
2. Department of Geography, Banaras Hindu University, Varanasi, Uttar Pradesh, INDIA
*kajal.kkm@gmail.com

Abstract
The analysis of flood frequency will depend on the historical peak discharge data for at least 10 years. This study has taken into account peak annual maximum discharge data for 72 years (1949 to 2020). The discharge data was collected from the Farakka Barrage Gauging station (24°48'15.10" N and 87°55'52.70" E) situated in the upper part of lower Ganga basin. The flood frequency analysis of the lower Ganga basin's upper portions has been carried out using Gumbel's frequency distribution method.

Gumbel's method (X_T) is a prediction analysing statistical approach. The discharge data was tabulated in descending order and rank has been assigned based on the discharge volume. The return period was calculated based on Weibull's formula (P) for this analysis. The flood frequency data was plotted on a graph where X-axis shows the return period and the Y-axis is the discharge value. The R^2 value of this graph is 0.9998 which describe Gumbel's distribution method is best for the flood frequency analysis. The flood frequency analysis is an essential step to assess the flood hazard.

Keywords: Flood Frequency, Lower Ganga Basin, Discharge data, Gumbel's method, Weibull's.

Introduction

The most frequent catastrophic event on the planet is flooding. Due to flood, alone, two third disaster-related damages have been observed worldwide[12]. In India, flood is a prevalent phenomenon. The major flood causing rivers of India's are Ganga, Brahmaputra, Godavari, Mahanadi etc. according to the RBA (Rashtriya Badh Ayog, India); at present approximate flood-prone area of India is 40 mHa (mega Hectare's) [3]. Flood causes a very devastating situation.

The present study mainly focuses on flood frequency analysis (FFA) based on the peak discharge data[4]. The flood frequency analysis is a historical data based probabilistic statistical approach[5]. Frequency analysis deals with the frequency of any data or group of data based on the central tendency[5]. The flood frequency analysis approach was constructed depending upon the annual maximum peak discharge data of particular/specific location or rivers[5,6]. Flood frequency analysis was introduced by the E.J. Gumbel, hence the name is Gumbel's probability distribution method[6].

Another most used method is Log-Pearson Type III (LP3) method[7,8]. In both methods, input variables are same (Peak Discharge data). Flood frequency analysis is an essential tool for analysing and prediction of floodplain or flood hazard area. It helps to reduce the flood related vulnerability and cost of the mitigation process[9]. It plays a vital role in providing estimates of the recurrence of floods which is helpful in designing major constructions such as bridges, dams, levees, drains, roadways, sewage disposal plants, waterworks and industrial structures[9]. The hydrologist or planner have used the flood frequency analysis technique to predict floods in specific recurrence period (time)[8]. Flood frequency analysis involves the probability model which describes the future predicted trend depending upon certain regions' previous periodic observation of river gauge station's discharge data[7].

The discharge data directly involves hydrological factors i.e. runoff, duration of rainfall and catchment area[6]. Flood frequency analysis provides a linear correlation with the hydrological data. Flood frequency analysis is very crucial for the present study because this study deals with the most flood prone river of India i.e. the Ganges river [3].

According to the historical data, flooding in Ganga river basin is very frequent, particularly in eastern Uttar Pradesh, entire Bihar and West Bengal (Maldah and Murshidabad district) States[10]. Every year during the monsoon season, due to the (rain) heavy water load on river bed, there has been frequent occurrence of flood on the low-lying catchment area. The main objective of the present study is to find out the flood frequency analysis for the lower Ganga River using the historical discharge data from Farakka gauging station which is near to the Maldah district of West Bengal.

Material and Methods

Study area: Ganga River is one of the major rivers in India which flows from North West to East. It has three divisions that are Upper Ganga (Gangotri to Allahabad), Middle Ganga (Allahabad to Rajmahal) and Lower Ganga Basin (Rajmahal to Bay of Bengal). The study area selected was Lower Ganga basin's upper part from Rajmahal (Jharkhand) to Farakka (Murshidabad). The geographical location of these areas is latitude 24°48' 00" to 25° 11'00 "N and longitude 87° 45' 00 "to 88°00' 00 "E. Discharge data for the present study was collected from the Farakka Barrage, which is located in the lower part of the study area. The geographical location of Farakka Barrage is 24°48'15.10" N (Latitude) and 87°55'52.70" E (Longitude) (Fig. 1).

For the present study, Farakka barrage discharge data was taken for the flood frequency analysis because the upstream part of Upper Lower Ganga basin's surrounding region (Maldah and Murshidabad district of West Bengal and Sahibganj District of Jharkhand) frequently suffers due to flood which severely impacts the surrounding population. Due to flood hazard in this study area, 1.6 lakh population (approx.) is directly or indirectly affected.[11-13] In this regard, to predict the future flood hazard or flood risk situation of this region, flood frequency analysis is essential.

Data Source: The flood frequency analysis is a multivariate statistical method for predicting the frequency of flooding. This is determined by the discharge information (Secondary). In this present study, 72 years annual peak discharge (1949 to 2020) data was collected from Farakka Barrage from three different sources (Table 1). The first source, 1949 to 1973 data was collected from GRDC (Global Runoff Data Centre, Koblenz, Germany) database[14], secondly , 1974 to 2003 was collected from Parua (1999-2003), Kalyan Rudra (2003), Bandyapadhy and Mallik Farakka Barrage Project[15] and thirdly, 2004 to 2020 data was collected from DFO (Dartmouth Flood Observatory) database (University of Colorado, Colorado, U.S.A.)[16]. The collected discharge data was tabulated together in a single table for analysing the flood frequency.

Gumbel's Methods: Gumbel's distribution is a probabilistic theory of statistics. It is used as a model of maximum number distribution among the various samples.[17] It is useful to predict the future natural disaster like flood, earthquake, drought etc.[9] The Gumbel's distribution method of frequency analysis needs minimum ten years annual maximum historical data to assume the probabilistic future prediction. It is also known as the generalized extreme value distribution method. In this study, the Gumbel frequency distribution method was applied to predict the flood frequency of lower Ganga River basin[3].

The 72 years annual maximum Peak discharge data was used to execute the flood frequency analysis. In flood frequency curve "X" axis represents the return period and "Y" axis represents annual maximum peak discharge value. The Gumbel's distribution flood frequency analysis was completed based on the equation number (1) and return period was calculated using the equation number (6).

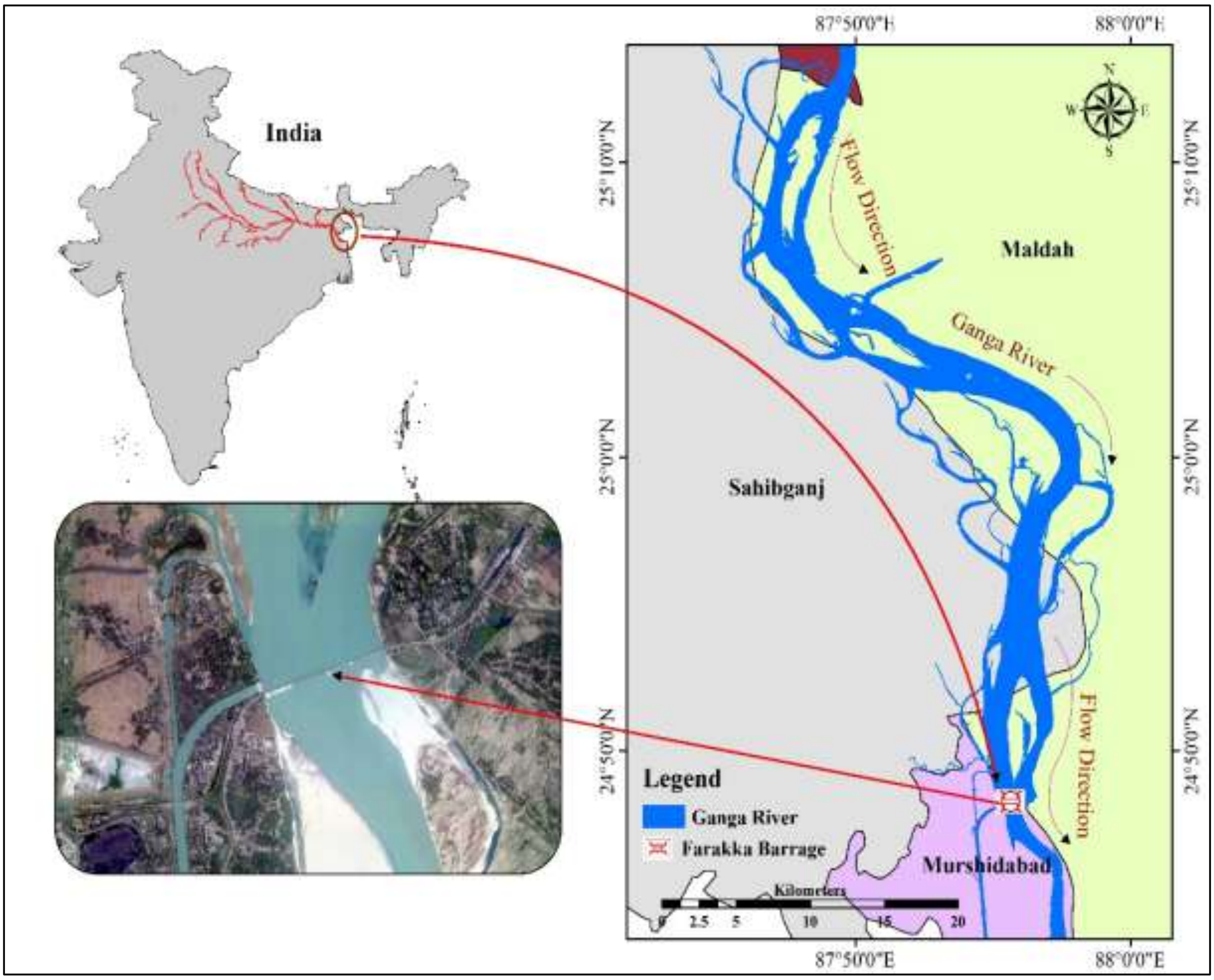

Figure. 1: Study area map

Table 1
Peak Discharge Data and its computation

Years	Discharge in m^3/s	Discharge in Descending order m^3/s (X)	Rank (M)	Return Period (P)	Probability Positions ($^1/_P$)	X^2
1949	40283	76830	1	73	0.013699	5902848900
1950	51340	74900	2	36	0.027778	5610010000
1951	31333	73200	3	23.66667	0.042254	5358240000
1952	40802	72379	4	17.5	0.057143	5238719641
1953	49411	72200	5	13.8	0.072464	5212840000
1954	62694	69900	6	11.33333	0.088235	4886010000
1955	61363	67400	7	9.571429	0.104478	4542760000
1956	47602	66900	8	8.25	0.121212	4475610000
1957	39765	66900	9	7.222222	0.138462	4475610000
1958	53410	65072	10	6.4	0.15625	4234365184
1959	43717	62694	11	5.727273	0.174603	3930537636
1960	40943	61363	12	5.166667	0.193548	3765417769
1961	40283	60500	13	4.692308	0.213115	3660250000
1962	51340	59900	14	4.285714	0.233333	3588010000
1963	31333	59000	15	3.933333	0.254237	3481000000
1964	40802	57660	16	3.625	0.275862	3324675600
1965	27810	57520	17	3.352941	0.298246	3308550400
1966	35570	56100	18	3.111111	0.321429	3147210000
1967	45930	55600	19	2.894737	0.345455	3091360000
1968	37104	54400	20	2.7	0.37037	2959360000
1969	46262	54300	21	2.52381	0.396226	2948490000
1970	38992	53500	22	2.363636	0.423077	2862250000
1971	65072	53410	23	2.217391	0.45098	2852628100
1972	24693	51340	24	2.083333	0.4800	2635795600
1973	44573	51340	25	1.96	0.510204	2635795600
1974	35570	50782	26	1.846154	0.541667	2578811524
1975	45930	49411	27	1.740741	0.574468	2441446921
1976	37104	48900	28	1.642857	0.608696	2391210000
1977	46262	48600	29	1.551724	0.644444	2361960000
1978	38992	48211	30	1.466667	0.681818	2324300521
1979	41100	47622	31	1.387097	0.72093	2267854884
1980	72200	47602	32	1.3125	0.761905	2265950404
1981	55600	46600	33	1.242424	0.804878	2171560000
1982	66900	46332	34	1.176471	0.8500	2146654224
1983	59900	46262	35	1.114286	0.897436	2140172644
1984	60500	46262	36	1.055556	0.947368	2140172644
1985	56100	45930	37	1.000	1.000	2109564900
1986	48600	45930	38	0.947368	1.055556	2109564900
1987	73200	45680	39	0.897436	1.114286	2086662400
1988	67400	45300	40	0.85	1.176471	2052090000
1989	35600	44573	41	0.804878	1.242424	1986752329
1990	54300	43717	42	0.761905	1.3125	1911176089
1991	59000	42545	43	0.72093	1.387097	1810077025
1992	45300	42186	44	0.681818	1.466667	1779658596
1993	53500	41521	45	0.644444	1.551724	1723993441
1994	66900	41280	46	0.608696	1.642857	1704038400
1995	48900	41100	47	0.574468	1.740741	1689210000
1996	69900	40943	48	0.541667	1.846154	1676329249
1997	46600	40802	49	0.510204	1.96	1664803204
1998	74900	40802	50	0.48	2.083333	1664803204
1999	57660	40673	51	0.45098	2.217391	1654292929

2000	76830	40283	52	0.423077	2.363636	1622720089
2001	45680	40283	53	0.396226	2.52381	1622720089
2002	54400	39765	54	0.37037	2.7	1581255225
2003	57520	39644	55	0.345455	2.894737	1571646736
2004	41280	38992	56	0.321429	3.111111	1520376064
2005	37437	38992	57	0.298246	3.352941	1520376064
2006	33512	37437	58	0.275862	3.625	1401528969
2007	48211	37104	59	0.254237	3.933333	1376706816
2008	42545	37104	60	0.233333	4.285714	1376706816
2009	29945	36188	61	0.213115	4.692308	1309571344
2010	50782	35600	62	0.193548	5.166667	1267360000
2011	46332	35570	63	0.174603	5.727273	1265224900
2012	36188	35570	64	0.15625	6.4	1265224900
2013	47622	33512	65	0.138462	7.222222	1123054144
2014	26876	32458	66	0.121212	8.25	1053521764
2015	39644	31333	67	0.104478	9.571429	981756889
2016	42186	31333	68	0.088235	11.33333	981756889
2017	32458	29945	69	0.072464	13.8	896703025
2018	40673	27810	70	0.057143	17.5	773396100
2019	72379	26876	71	0.042254	23.66667	722319376
2020	41521	24693	72	0.027778	36	609744249

The Gumbel's Distribution time (T) dependent probability frequency analysis equation is (1):

$$X_T = \overline{X} + K.\sigma_X \qquad (1)$$

where X_T is Gumbel's Distribution in reference to return period; $\overline{X}$ is the mean value; σ_X is the standard deviation; and "K" is the factor of frequency in Gumbel method.

The mean value and σ_X are derived from the equation (2 and 3):

$$\overline{\overline{X}} = \frac{\sum X}{N} \qquad (2)$$

where "X" is the discharge value, $\overline{\overline{X}}$ is the mean of the discharge and "N" is the number of samples.

$$\sigma = \sqrt{\frac{\sum_{(i-1)}^{n}(Xi - \overline{\overline{X}})^2}{n}} \qquad (3)$$

where σ = standard deviation, "n" is the number of sample, " Xi " is the each value of the sample and $\overline{\overline{X}}$ is the mean value of this sample.

The "K" value was calculated using the following equation (4):

$$K = \frac{Y_T - \overline{\overline{Y_n}}}{S_n} \qquad (4)$$

where Y_T is the reduced variate which is calculated by using the equation (5); the S_n and $\overline{\overline{Y_n}}$ value have been used from Gumbel's extreme value distribution chart that depends on the sample size.

$$Y_T = -\left[\text{Ln.Ln.}\left(\frac{T}{T-1}\right)\right] \qquad (5)$$

where "T" is the predicted time period.

$$P = \frac{(m-a)}{(N-a-b+1)} \qquad (6)$$

where "P" is the plotting position, "m" is the rank, "N" is the lowest order of the sample and "a and b" are the constant value.

Methodology

The flood frequency analysis (FRA) was carried out using Gumbel's frequency distribution method. During this study, the collected discharge data was arranged in descending order (X). The rank (M) was assigned based on the discharge volume (Table 1)[9]. The first (1) rank was assigned for the highest discharge value i.e. 76,830 m^3/s and the lowest rank (72) was assigned for the lowest discharge value i.e. 24,693 m^3/s[9,18]. The return period (P) was calculated based on Weibull's equation (6) and FRA (X_T) was calculated by applying equation (1). To complete the X_T (FRA) and $\overline{X}$ (Mean), σ (Standard deviation) was calculated from the peak discharge data using the equations (2 and 3).

The "K" value of this study was calculated based on equation (4). In the "K" value equation S_n and $\overline{\overline{Y_n}}$ are the constant values taken from the Gumbel's distribution chart (sample wise constants were obtained)[18]. "Y_T" value was generated using the equation (5) which was used for calculating the K value[19]. After that, all computed data was used to generate the flood frequency of this study area.

Discussion

The flood frequency analysis was carried out by adapting the above methodology. In this study area, flood is a normal phenomenon occurring during the rainy seasons (July to September). 72 years historical data (1949 – 2020) was used to complete this study (Table 1). From the tabulated data, it was observed that the highest peak discharge value was 76,830 m^3/s (2000) and the lowest peak discharge value was 24,693 m^3/s (1971). The peak Discharge data was grouped into two classes (<50000 m^3/s and >50000 m^3/s). This classification is completely based on the leaner method (Fig. 2) where R^2 value is 0.0057, which is almost a straight line. From this discharge data table, it was observed that 46 years of data come under <50000 m^3/s and 26 years of data come under >50000 m^3/s.

The maximum return period (P) value of this analysis was 73 and the minimum value was 0.0277 which was calculated based on Weibull's return period equation (6) (Table 1). The probability positions value was calculated (Table 1) by inversing the return period ($^1/_p$). The highest $^1/_p$ value was 36 and the lowest was 0.013699. The probability positions data and discharge data were plotted on figure 2, which shows that the high discharged return probability is less and the low discharged return probability is very high (Fig. 3). For the present data, mean discharge value was 47,977.30 m^3/s and the standard deviation value was 12,500.11 (equation 2 and 3). These values were used to calculate the Y_T, K and final X_T (flood frequency) for this study. The Y_T (equation 5), K (equation 4) and X_T (equation 1) were separately calculated depending upon the $\overline{X}$ and σ. In this study, the FRA's return period was taken for 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250 and 300 years. According to the X_T equation, the flood frequency discharge value was generated (Table 2) for these study areas.

Lastly, the T and X_T value were plotted on the graph (Fig. 4) where R^2 = 0.9998 [9]. The R^2 value of this scatter plot has justified that Gumbel's statistical distribution method is suitable for analysing the flood frequency. The X-axis of this graph shows the return period and the Y-axis is the discharge value (Fig. 4).

Conclusion

The flood frequency analysis is a very realistic approach to warn and to predict the long term and near feature discharge flow character of a river. In the present study, flood frequency analysis has been completed for the lower Ganga River (Farakka Barrage) using 72 years peak historical discharge data (Fig. 2). In this study, Gumball's Flood Frequency Analysis has been used; this is the most popular method to analyse the flood frequency. Flood frequency curve shows that the high discharge occurrence depends upon the long-time interval, but the low discharge frequently repeats in this study area (Table 2 and fig. 4).

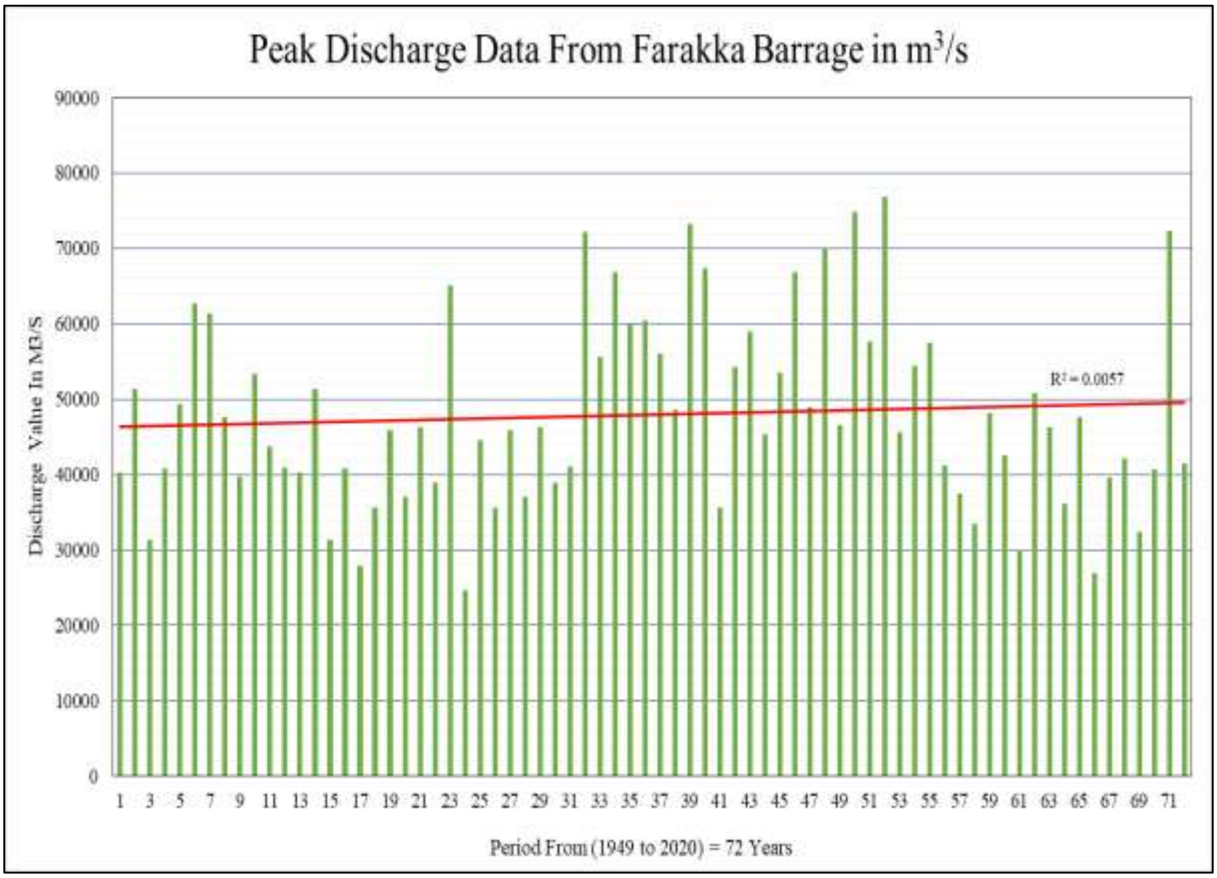

Figure 2: Yearly Peak Discharge Data from 1949 – 2020

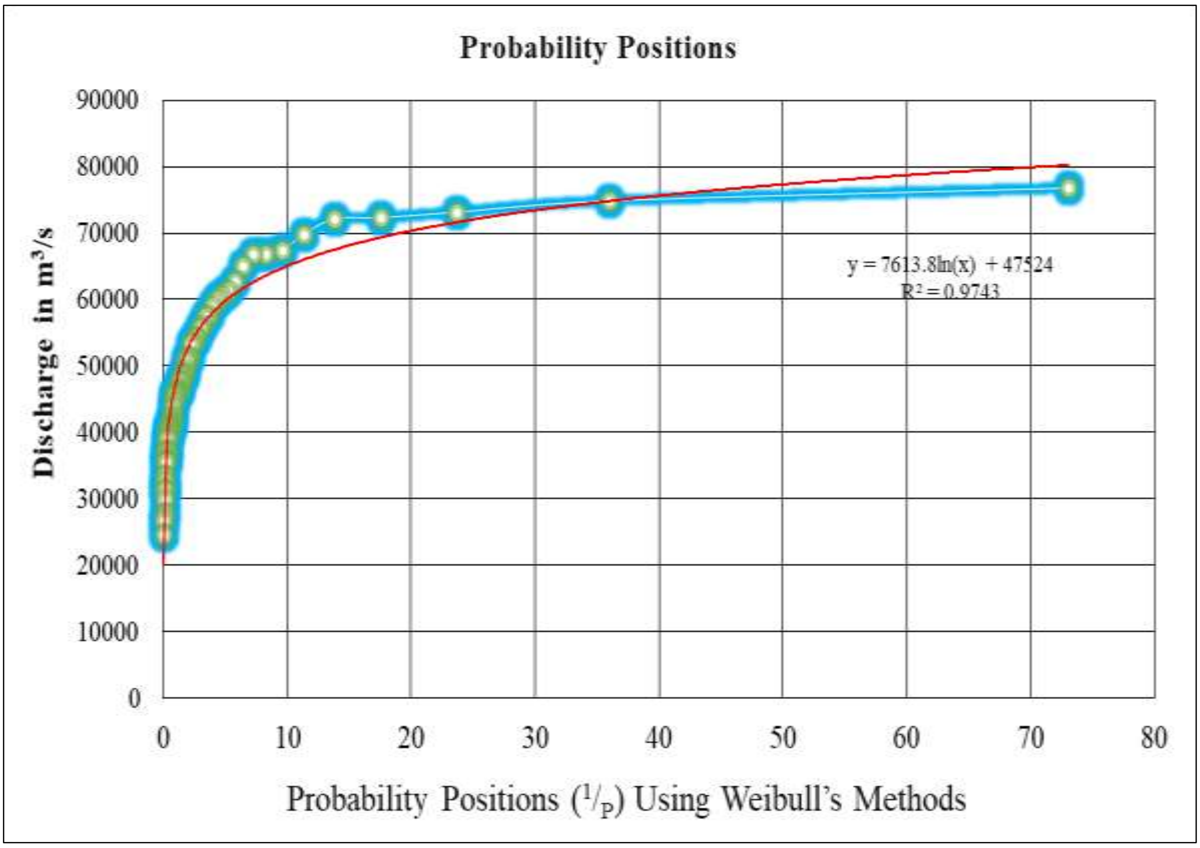

Figure 3: Probability Positions using Weibull's method.

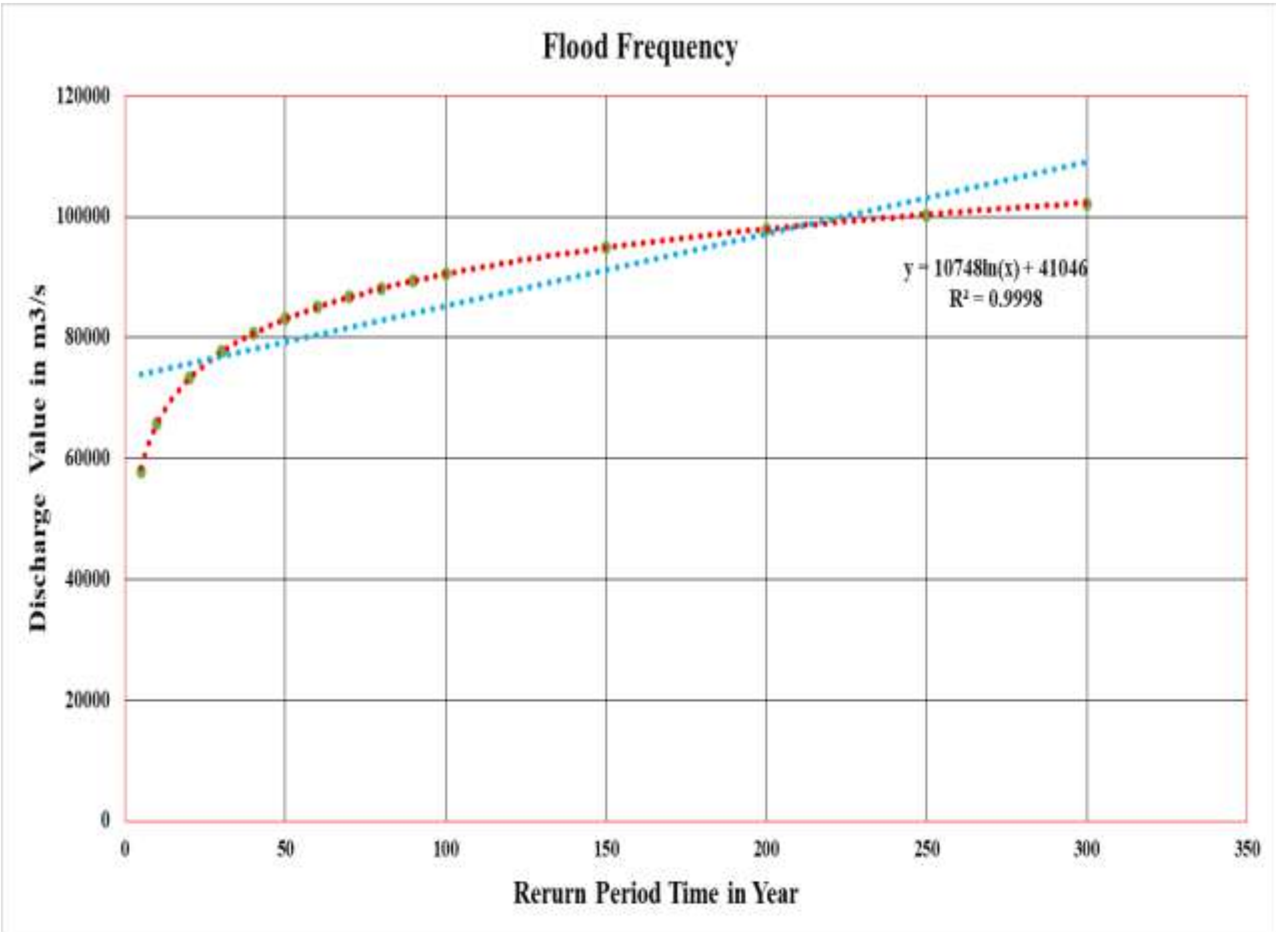

Figure 4: Flood Frequency Analysis Graph

Table 2
Calculation of Discharge value using different (T) Return Periods

Return Period In Year (T)	Mean/ $\overline{X}$	σ	Y_T	K	X_T (m^3/s)
5	47977.30	12500.11	1.499939987	0.795704529	57923.702
10	47977.30	12500.11	2.250367327	1.427749791	65824.340
20	47977.30	12500.11	2.970195249	2.034022782	73402.822
30	47977.30	12500.11	3.384294493	2.382796676	77762.535
40	47977.30	12500.11	3.676247258	2.62869305	80836.268
50	47977.30	12500.11	3.901938658	2.81878098	83212.388
60	47977.30	12500.11	4.085952773	2.973766338	85149.723
70	47977.30	12500.11	4.2413095	3.104615093	86785.347
80	47977.30	12500.11	4.375743836	3.217842025	88200.697
90	47977.30	12500.11	4.494228222	3.317635157	89448.123
100	47977.30	12500.11	4.600149227	3.406846818	90563.278
150	47977.30	12500.11	5.007292664	3.749762204	94849.760
200	47977.30	12500.11	5.295812143	3.992766902	97887.346
250	47977.30	12500.11	5.519457577	4.181131623	100241.927
300	47977.30	12500.11	5.702113489	4.334973038	102164.962

FRA is essential for this study area because this area is highly flood-prone. Due to flood every year approx. 2 lakh population is severely affected. This study will help the disaster management authority to pre-mitigate the flood-prone area (particularly the Maldah and Murshidabad districts of West Bengal) and reduce flood-related hazards.

References

1. Benameur S. et al, Complete flood frequency analysis in Abiod watershed, Biskra (Algeria), *Natural Hazards*, **86(2)**, 519–534, doi: 10.1007/s11069-016-2703-4 **(2017)**

2. Bhagat N., Flood Frequency Analysis Using Gumbel's Distribution Method: A Case Study of Lower Mahi Basin, India, *Journal of Water Resources and Ocean Science*, **6(4)**, 51, doi: 10.11648/j.wros.20170604.11 **(2017)**

3. Bhat M.S. et al, Flood frequency analysis of river Jhelum in Kashmir basin, *Quaternary International*, **507**, 288–294, doi: 10.1016/j.quaint.2018.09.039 **(2019)**

4. Burn D.H., Evaluation of regional flood frequency analysis with a region of influence approach, *Water Resources Research*, **26(10)**, 2257–2265, doi: 10.1029/WR026i010p02257 **(1990)**

5. DFO, The Flood Observatory, INSTAAR, University of Colorado, Available at: https://floodobservatory.colorado.edu/ Accessed: 25 March 2021 **(2020)**

6. Ganamala K. and Sundar Kumar P., A case study on flood frequency analysis, *International Journal of Civil Engineering and Technology*, **8(4)**, 1762–1767 **(2017)**

7. GRDC, BfG - GRDC Data Download, Available at: https://www.bafg.de/GRDC/EN/02_srvcs/21_tmsrs/210_prtl/prtl_node.html, Accessed: 25 March 2021 **(1990)**

8. Hallegatte S. et al, Flood Risks, Climate Change Impacts and Adaptation Benefits in Mumbai, An Initial Assessment of Socio-Economic Consequences of Present and Climate Change Induced Flood Risks and of Possible Adaptation Options, Available at: www.oecd.org/env/workingpapers, Accessed: 25 March 2021 **(2010)**

9. Hosking J.R.M. and Wallis J.R., The Value of Historical Data in Flood Frequency Analysis, *Water Resources Research*, **22(11)**, 1606–1612, doi: 10.1029/WR022i011p01606 **(1986)**

10. Islam N., Flood Risk assessment of Murshidabad District in West Bengal Using Remote Sensing and GIS Techniques, Pondicherry University **(2017)**

11. Kaus S., Disaster Management Plan 2019-2020, Malda District **(2020)**

12. Kumar R., Flood Frequency Analysis of the Rapti River Basin using Log Pearson Type-III and Gumbel Extreme Value-1 Methods, *Journal of the Geological Society of India*, **94(5)**, 480–484, doi: 10.1007/s12594-019-1344-0 **(2019)**

13. Mandal K.K. et al, Investigation of Cropping Pattern of Jalangi Block, West Bengal, *IOSR Journal Of Humanities And Social Science (IOSR-JHSS)*, **25(4)**, 58-64, doi: 10.9790/0837-2504015864 **(2020)**

14. Pal S.C. et al, Flood Frequency Analysis and Its Management in Selected Part of Bardhaman District, West Bengal, *Advances in Asian Human-Environmental Research*, 225–246, doi: 10.1007/978-3-030-49115-4_13 **(2021)**

15. Rabolin C., India Water Resources Information System (IWRIS), Ganga Basin, Available at: https://indiawris.gov.in/wris/#/Basin, 1-7 **(2011)**

16. Ritchie H., Natural Disasters, in Our World in Data, Our World in Data, Available at: https://ourworldindata.org/natural-disasters#citation, Accessed: 25 March 2021 **(2014)**

17. Singo L.R. et al, Flood frequency analysis of annual maximum stream flows for Luvuvhu River Catchment, Limpopo Province, South Africa, 16th SANCIAHS Hydrology Symposium, 1–3 **(2012)**

18. WHO, Floods, Available at: https://www.who.int/health-topics/floods#tab=tab_1, Accessed: 25 March 2021 **(2016)**

19. Yue S. et al, The Gumbel mixed model for flood frequency analysis, *Journal of Hydrology*, **226(1–2)**, 88–100, doi: 10.1016/S0022-1694(99)00168-7 **(1999)**.

(Received 30th March 2021, accepted 06th June 2021)