Journal of Civil Engineering (IEB), 33 (1) (2005) 9-17

Journal of Civil Engineering IEB

Study of plotting position formulae for Surma basin in Bangladesh

M. J. B. Alam[1] and A. Matin[2]

[1]*Department of Civil and Environmental Engineering*
Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh
[2]*Department of Water Resources Engineering*
Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh

Received on 05 January 2004

Abstract

Selection of an appropriate frequency distribution is of utmost importance in flood frequency analysis. A particular distribution is judged on the basis of its ability to fit the observed data for a river basin. Plotting position of observed data series is ascertained with some empirical or semi-empirical plotting position formula. Fitting of a particular data series to a distribution depends on the formula for plotting position being used. Existing practice in selection of a particular formula is arbitrary and often gives unrealistic results. Investigation has been carried out to determine the most suitable plotting position formula for the Gumbel distribution. On the basis of some statistical criteria, 11 plotting position formulae have been evaluated with annual flood series of 21 rivers of the Surma basin having record length from 10 to 37 years. Weibull plotting position formula has been found to fit best for the Gumbel distribution followed by the Adamowski formula.

Keywords: Surma basin, frequency distribution, Gumbel method, plotting position

1. Introduction

One of the major problems faced in water resource engineering is the estimation of design flood. There is no means of forecasting the exact sequence of future flood events at any location. However if it is assumed that the sequences which will occur have the same statistical characteristics as sequences that had occurred in the past, then it is possible to estimate the probability of any magnitude being exceeded during the design life of some schemes. Prediction of flood by statistical means deals with the analysis of flood data at a given site, which exceeded on an average once in certain years of return period. Real time flood forecasting is an important non-structural measure for reducing flood damages.

Prediction of flood is required in the design of dams, highway and railway bridges, culverts, flood control structures, etc. The American Water Works Association had reported that out of 293 dam failures in the U.S. and other countries since 1800, about 20% were due to faulty spillway design. This only highlights the importance of an accurate flood prediction. Due to inaccurate method of prediction, there may be dam failure, highway damage, loss of human and animal lives and loss of property.

Various investigators suggested different methods for estimating and predicting design floods. Statistical method of frequency analysis is one of such methods for estimating design flood from recorded flood events. In frequency analysis sample data series is used to fit frequency distribution, which in turn is used to extrapolate from recorded events to design events either graphically or by estimating the parameters of frequency distribution. The graphical analysis consists of assigning a probability of exceedence or non-exceedence to each ordered observation on the basis of appropriate probability paper. The probability of exceedence or non-exceedence is attached to each observation on the basis of appropriate plotting position formula.

Many probability distributions and various ways of fitting the data series are available. Inspite of significant developments, the selection of an appropriate distribution for any given flood record from among the alternate distributions is still a subject of continuing investigation. Probability plots are used in frequency analysis to fit the probability distributions to a given series to identify the outliers and to assess goodness of fit. Recorded flood events are plotted on probability paper or semi-log paper with magnitude of flood against return period of a probability of exceedence along with the theoretical distribution to test for the goodness of fit. The return period or probability of exceedence of any event of the recorded series is computed on the basis of certain assumed plotting position. There are a number of plotting position formulae put forward by various investigators with their own justifications. There is wide variation of the return period or probability of exceedence for a particular event in a series, when computed with different formulae, especially when the event lies in the extremities of the series. Thus, the goodness of fit for a statistical distribution depends on the plotting position formula adopted. This necessitates the study for determining the most suitable plotting position formula for a given distribution. Extreme value type1 (EV-1) distribution, introduced by Gumbel (1958), is most often used for frequency analysis of extremes in meteorology and hydrology. The Brahmaputra Flood Control Commission (1977) has also preferred the Gumbel distribution for flood frequency analysis. The present study is restricted to Gumbel distribution only.

2. Plotting position formulae

The purpose of the frequency analysis of an annual series is to obtain a relationship between the magnitude of an event and its probability of exceedence. The probability analysis may be made either by empirical or by analytical methods. A simple empirical technique is to arrange the given annual extreme series in descending order of magnitude and to assign an order number called rank to the event value. The probability P of an event equaled to or exceeded is then calculated by an empirical or semi-empirical formula and there are numerous such formulae available to calculate P. The exceedence probability of the event obtained by the use of an empirical formula is called its plotting position. Table 1 shows the various plotting position formulae and their range in recurrence intervals obtained for 10 years record length.

Chow (1964) has demonstrated theoretically that Callifornia method is suitable for plotting annual exceedence series or partial duration series. However, this simple formula plots the data at the edge of group intervals and also produces a probability of 100 percent for m=N, which may not be easily plotted on a probability scale. Hazen (1930) proposed his formulae, which plots data at the centers of group intervals. As the extreme distribution was later introduced to frequency analysis, the Weibull (1939) formula was soon found to be a compromise with more statistical justification. Chow (1964) has shown that this formula is theoretically suitable for plotting the annual maximum series; the US Water Resources Council adopted this formula as the standard plotting position method.

Another compromise is the Chegodajew formula widely used in Russia and the Eastern European countries. Cunnane (1978) studied the various available plotting position methods using criteria of unbiasedness and minimum variance. For normally distributed data, he found that the Bloom (1958) plotting position is close to being unbiased, while for data distributed according to the EV-1 distribution, the Gringorton (1963) formula is the best. Adamwaski (1981) advocates for his formula, for EV-1 distribution, for high values of probability of exceedence based on mean square criteria. Similarly, other plotting position formulae have been recommended for one reason or the other.

All the methods of determining plotting positions give practically the same results in the middle of a distribution but produces different positions near the tails of the distribution. Thus the choice of a plotting position formula becomes important for fitting the extreme value flood data.

Table 1
Various plotting position formulae

Sl. No.	Method	Formula P (X > x)	Range of Return Period (Years)
1	Callifornia	m / N	10.0
2	Hazen	(m - 0.5) / N	20.0
3	Weibull	m / (N + 1)	11.0
4	Beard	(m - 0.31) / (N + 0.38)	15.0
5	Benard	(m - 0.3) / (N + 0.2)	14.7
6	Chegodajew	(m - 0.3) / (N + 0.4)	14.9
7	Blom	(m – 3 / 8) / (N + 1 / 4)	16.4
8	Tukey	(m – 1 / 3) / (N + 1 / 3)	15.4
9	Gringorton	(m - 0.44) / (N + 0.12)	18.1
10	Cunnane	(m - 0.40) / (N + 0.2)	17.2
11	Adamowski	(m – 0.25) / (N + 0.5)	14.0

The Surma basin comprises an area of 24,265 km2 and constitutes 15% of the country. It is a large, gentle depression feature, bounded by the old Brahmoputra flood plain in the west, the Shillong plateau’s foothills in the north and by the Sylhet high plain in the east. Its greatest length, both E-W and N-S, is just over 113 km. Numerous lakes (beels) and large hoars cover this saucer-shaped area of about 7,250 sq.km. The land of the region and its adjacent tributary areas play an important role in determining the spatial distributions of the rainfall, surface and groundwaters within the region (Ali and Amin, 2005).

3. Methodology

In order to compare different plotting position formulae for Gumbel distribution, the expected peak discharges for different return period is calculated by a particular plotting position formula and are compared with the observed peak discharge for that return period obtained for Gumbel distribution. The comparison is made on the basis of overall fit, fit in the upper tail region and fit in the lower tail region.

Overall fit

$$SSE = \sum_{i=1}^{N} \left[Q(i)obs. - Q(i)theo. \right]^2 \tag{1}$$

The overall fit is judged on the basis of total sum of square of error (SSE)

Fit in the Upper Tail Region

The following two criteria has been used to judge the performance of fit in the upper tail region:

(a) Sum of square of error in top 3 peak discharges (SSET3)

$$SSET3 = \sum_{i=1}^{3} \left[Q(i)obs. - Q(i)theo. \right]^2 \tag{2}$$

(b) Sum of square of error in top 6 peak discharges (SSET6)

$$SSET6 = \sum_{i=1}^{6} \left[Q(i)obs. - Q(i)theo. \right]^2 \tag{3}$$

Fit in the Lower Tail Region

The following two criteria has been used to judge the performance of fit in the lower tail region:

(a) Sum of square of error in bottom 3 peak discharges (SSEB3)

$$SSEB3 = \sum_{i=N-2}^{N} \left[Q(i)obs. - Q(i)theo. \right]^2 \tag{4}$$

(b) Sum of square of error in bottom 6 peak discharges (SSEB6)

$$SSEB6 = \sum_{i=N-5}^{N} \left[Q(i)obs. - Q(i)theo. \right]^2 \tag{5}$$

The return periods (T) of observed discharge for different plotting position formulae are calculated by using following generalized formula:

$$T = \frac{N + B}{m - A} \tag{6}$$

where, A and B are constants whose value depends on the type of plotting position formula.

The expected discharges for the Gumbel distribution for different return periods are calculated by the following formula:

$$Qt = M + K \times SD \tag{7}$$

where, Qt= expected discharge with recurrence interval T
M=mean of the observed data
K= frequency factor expressed as

$$K = \frac{Yt - Yn}{Sn} \tag{8}$$

where, Yt= reduced variate, a function of T and is given by

$$Yt = -\left[\ln . \ln \frac{T}{T-1}\right] \tag{9}$$

Yn = reduced mean, a function of sample size N
Sn = reduced standard deviation, a function of sample size N
SD = standard deviation

4. Results and discussion

The river under the study was Surma along with its twenty major tributaries on the north and south bank basin. The length of recorded discharge varies from 10 years to 37 years. Eleven plotting position formulae, as shown in the Table 1, were compared. To compare the different plotting position formulae, the summary of the complete results for five rivers, out of twenty-one, are shown in the Tables 2 to 6. It is observed from the tables that best fitted plotting position formula according to the criteria discussed earlier is Weibull. The observed flood peaks for river Surma are fitted against Gumbel distribution in Fig.1 according to different plotting position formulae. The suitability of the different plotting position formulae for EV-1 in flood frequency analysis are discussed below.

4.1 Overall Fit

For almost all the data set of the rivers, Weibull plotting position formula gives minimum sum of squares of error of observed discharge and theoretical discharge. The sum of squares of error obtained by Adamowaski formula gives close result to that of Weibull formula. Chegodajew and Tukey plotting position formulae are better than others, except Weibull and Adamowski.

4.2 Fit in the Upper Tail Region

There is no consistent result when sum of squares of error in top 3 observed and theoretical discharges are considered. But Cunnane, Weibull and Callifornia plotting position formulae are found to be better than the otherformulae. Weibull plotting position formula gives minimum sum of squares of error in top 6 observed and theoretical discharges Adamowski and California formulae are better than other formulae. From Fig.1 it is seen that Weibull formula gives more close theoretical discharge with the observed data than other formulae for upper most tail region.

4.3 Fit in the Lower Tail Region

From the tables it is found that Weibull plotting position formula gives minimum sum of squares of error for both bottom 3 and bottom 6 observed and theoretical discharges for most of the rivers. Adamowski and Benard plotting position formulae are better than other formulae. From Fig. 1 it is found that Weibull plotting position formula is better fitted in the lower and upper tail region than otherfformulae. Data has been collected from the Bangladesh Water Development Board (BWDB) and Institute of Water Modeling (Ehsan and Suman, 2003).

Table 2
Comparison of different plotting position formulae for the river Surma

Plotting position Formula	SSE	SSET3	SSET6	SSEB3	SSEB6
California	134802900	25386900	31603810	18084960	18513850
Hazen	167326200	5588700	89360960	2347273	2871377
Weibul	117869300	27611340	34931740	8500937	9737958
Beard	141386000	50558930	61972510	4818757	5612835
Chegodajew	136120500	45915720	56658480	4543129	5296145
Gringorton	155426400	64312630	7710602	2946237	3530378
Blom	145222000	54618170	66420130	3658454	4316125
Tukey	139856200	49497700	60700280	4142591	4851864
Benard	134240500	45006950	55381490	2278049	2774220
Cunnane	148859400	58078560	70253580	3377748	4006111
Adamowski	131321400	41280780	51365500	5162615	5985196

Table 3
Comparison of different plotting position formulae for the river Manu

Plotting position Formula	SSE	SSET3	SSET6	SSEB3	SSEB6
California	29708150	3498948	5542798	19472850	22418480
Hazen	20482870	4980243	9040324	7037121	8976440
Weibul	13897490	3621744	6236789	4384194	5791305
Beard	16565120	3367847	6813228	5811828	7531865
Chegodajew	16423440	3334637	6750063	5756511	7465620
Gringorton	18944630	4225614	8081917	6609429	8477054
Blom	17618180	3687327	7333104	6191156	7983694
Tukey	16916170	3458936	6973354	5948090	7694875
Benard	17734500	3239589	6568410	7103827	9100448
Cunnane	18092160	3865042	7590568	6347032	8168141
Adamowski	15785280	3222626	6491616	5490904	7146235

Table 4
Comparison of different plotting position formulae for the river Dhahli

Plotting position Formula	SSE	SSET3	SSET6	SSEB3	SSEB6
California	2023591	442973	867971	851316	1139308
Hazen	2286216	1650797	1976443	119253	267431
Weibul	118130	477089	845735	124868	232425
Beard	1614545	991591	1332774	109167	240391
Chegodajew	1589622	966698	130873	109173	239560
Gringorton	2025445	1396768	172718	11362	256229
Blom	1797890	1173405	1509137	110257	247053
Tukey	1675046	1051669	1390933	109344	242662
Benard	1608580	941878	1290088	122999	284867
Cunnane	1879554	1253799	1587471	111256	250257
Adamowski	1476334	852820	1199142	109820	236099

Table 5
Comparison of different plotting position formulae for the river Khaiy

Plotting position Formula	SSE	SSET3	SSET6	SSEB3	SSEB6
California	249487	21113	30713	164234	175825
Hazen	151459	23348	47080	51884	58210
Weibul	112710	20210	32346	51884	58210
Beard	1221136	10531	29169	41425	46572
Chegodajew	121262	10407	28801	40957	46047
Gringorton	139046	16684	38705	48212	54147
Blom	129181	12443	32720	44643	47865
Tukey	124390	10978	30180	52278	47865
Benard	131466	10112	27759	52278	58813
Cunnane	132606	13762	34696	45970	51654
Adamowski	117618	10356	27570	38718	43527

5. Conclusions

On the basis of different statistical criteria it is found that Weibull plotting position formula is best fitted for the Gumbel distribution for rivers of the Surma basin. Adamowski formula is better than other formulae. Weibull plotting position formula gives minimum sum of squares of error of observed and theoretical discharges. This is a strong statistical criterion as the magnitude of individual observed data can be taken into account. Weibull formula also fits better in the extremities of the series. Neglecting some exceptional cases, which may be due to the error in the recorded data, or due to the limited number of recorded data, it may be concluded that Weibull plotting position formula is better fitted for Gumbel distribution for the rivers of Surma basin.

Table 6
Comparison of different plotting position formulae for the river Khishyara

Plotting position Formula	SSE	SSET3	SSET6	SSEB3	SSEB6
California	296714	20749	38404	183530	217477
Hazen	179735	35693	64136	37667	58485
Weibul	117187	15316	37259	15320	30171
Beard	136333	11792	37243	26606	44950
Chegodajew	134956	11297	36610	26134	44355
Gringorton	161773	24515	51945	33681	53688
Blom	147133	16536	42931	29910	49070
Tukey	139822	13153	38922	27779	46425
Benard	146690	11345	35379	38192	59929
Cunnane	152253	19171	45950	31300	50782
Adamowski	129042	9622	34271	23903	41521

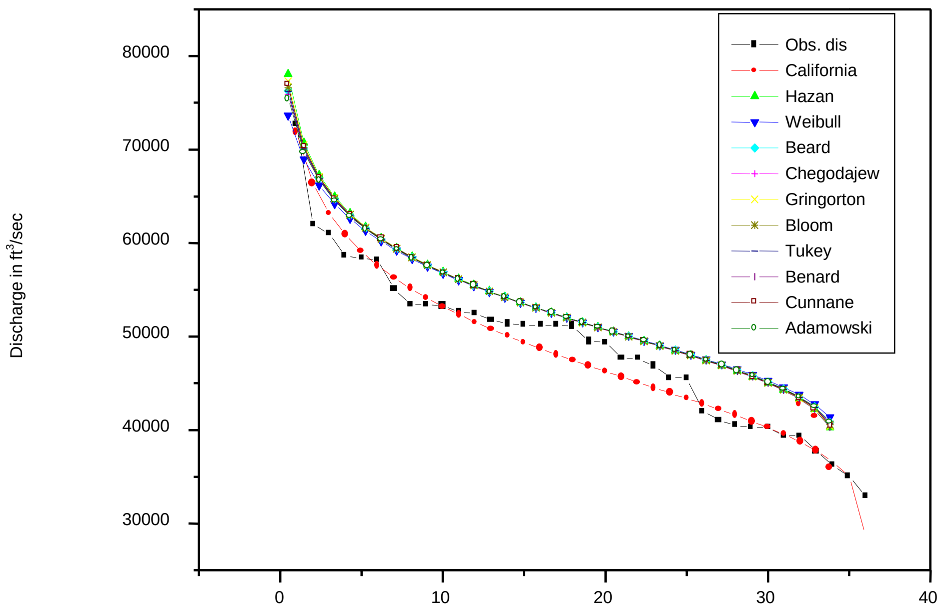

Fig. 1. Observed discharge versus theoretical discharge calculated using different plotting position formulae for the river Surma

References

Ali, M. and Amin, M. M. (2005). "Study of river system and flow routing of north east region of Bangladesh", B.Sc Engineering Thesis, Civil and Environmental Engineering Department, SUST, Sylhet, Bangladesh.

Adamowski, K. (1981). Plotting Formulae for Flood Frequency, *Water Resource Bulletin*, 17(2), pp. 197.

Brahmaputra Flood Control Commission, (1977). Comprehensive Plan for Brahmaputra River, 1, pp.243-246.

Bloom, G. (1958). "Statistical Estimates and Transformed Beta Variables", Wiley, New York, pp.197.

Chow, V.T. (1964). Hand Book of Applied Hydrology, McGraw-Hill, New York.
Cunnane, C. (1978). Unbiased Plotting Position:A Review, J. of Hydrology, 37, pp. 205-222.
Ehsan, K. Haque and Suman, K. Nath (2003). Hydrological Morphology Study of Rivers of Northern Region of Bangladesh, Internee Report, IWM.
Gringorton, I.I. (1963). A plotting Rule for Extreme Probability Paper, J. of Geophysical Res., 68(3), pp.813.
Gumbel, E.J. (1958), Statistics of Extremes. Colombia University Press, New York.
Hazen, A. (1930). Flood Flows-A Studies of Frequencies and Magnitudes, John Wiley and Sons Inc., New York.
Weibull, W. (1939). A Statistical Theory of Strength of Materials. Stockholm, pp.151.