TR-9

SOME STUDIES ON PLOTTING POSITION FORMULAE FOR GUMBELL EV-1 DISTRIBUTION

SATISH CHANDRA

DIRECTOR

STUDY GROUP

S M SETH

N K GOEL

NATIONAL INSTITUTE OF HYDROLOGY

JAL VIGYAN BHAVAN

ROORKEE - 247667 (U.P.)

INDIA

1986-87

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Probability plots are used in flood frequency analysis to fit the probability distributions to given flood series, to identify the outliers and to assess goodness of fit. If the objective of such a plot is to determine distribution parameters, then one must use unbiassed plotting positions.

Gumbel extreme value type 1 distribution is one of the commonly used distributions in India for flood frequency analysis. At present many plotting position formulae are in use for Gumbel EV-1 distribution. These formulae provide different results, particularly at the tails of the distribution.

In the present study various plotting position formulae have been compared with the unbiassed plotting positions (the expected value of reduced variates) on the basis of seven statistical criteria for Gumbel EV 1 distribution. Unbiassed plotting positions have been obtained from synethetically generated EV 1 reduced variates for different sample sizes. The results indicate that the plotting positions given by Gringorton formula $P(X \geqslant x) = (i - 0.44)/(N + 0.12)$ are more close to the unbiassed plotting positions (expected value of reduced variates) for all the sample sizes (10-100) on the basis of following four criteria

i) total sum of squares of error,

ii) absolute error in highest reduced variate,

iii) sum of squares of error in top 3 reduced variates and

iv) sum of squares of error in top 6 reduced variates

2. The performance of cunnane and Hazen plotting position formulae is better than all other formulae (except Gringorton) on the basis of above 4 criteria.

3. The plotting positions given by Tukey formulae $P(X \geqslant x) = (i - 1/3)/(N+1/3)$ are more close to the unbiassed plotting positions in the lower tail, on the basis of

   i) absolute error in lowest reduced variate,

   ii) sum of squares of error in bottom 3 reduced variates and

   iii) sum of squares of error in bottom 6 reduced variates.

This clearly indicates the suitability of Gringorton plotting position formula for flood frequency analysis using Gumbel EV-1 distribution, which involves extrapolation in the upper tail region.

1.0 INTRODUCTION

Water resources engineers are generally concerned with the amount of water, known as design flood, a hydraulic structure is going to face during its economic life. The hydraulic structure may be weir, barrage, bridge or dam etc. Generally the following three approaches are used to estimate the design flood :

i. Empirical formulae,

ii. Deterministic approach and

iii. Statistical approach,

1.1 Empirical Formulae

The earliest approaches to the estimation of future floods on a drainage basin were based upon simple empirical formulae involving the correlation of past peak discharges with various basin parameters like, area, width and length of drainage basin, etc.

Formulae for the maximum expected flood involving drainage area only are of the general type

$$Q = C A^n \qquad \ldots (1)$$

Where C is a constant depending upon the characteristics of the drainage basin, A is the drainage area in square miles, n is a constant varying from 0.5 to 1.0 and Q is the flood flow rate in cusecs.

Dickens, Ryves, Inglis, Fanning, Charmier, Craig and Rhind etc. are some of the important empirical formulae. Varshney (1979) gives extensive list of empirical formulae, developed for Indian catchments. The inconsistent results from the application of such empirical formulae have made their use very limited.

1.2 Deterministic Approach

Deterministic approach for design flood estimation involves estimation of design storm and derivation of unit hydrograph. The unit hydrograph can be used to estimate design flood from design storm.

In case of big basins, the basin is divided into various sub basins. The flood for various sub basins are estimated using unit hydrograph and routed to the destination. Muskingum method or Kalinin Miliyukov method may be used for routing the flood.

1.3 Statistical approach

In statistical approach or frequency analysis approach the sample data is used to fit frequency distribution which in turn is used to extrapolate from recorded events to design events either analytically or graphically.

1.3.1 Analytical approach

In analytical approach the T year flood is obtained using the following formula

$$Q_T = \bar{Q} + K_T \cdot S \qquad \ldots (2)$$

Where

Q_T = T year flood,

$\bar{Q}$ = Mean of the sample (annual peak discharge series)

S = Standard deviation of the sample and

K_T = Frequency factor (which depends upon the probability distribution and the return period)

1.3.2 Graphical approach

The graphical analysis consists of assigning a probability of exceedence or non exceedence to each ordered observation and selection of an appropriate probability paper. The probability of exceedence or non exceedence is attached to each observation on the basis of plotting formula.

Many distributions and various ways of fitting them are available (Yevjevich, 1972). The selection of an appropriate distribution for any given flood records from among the alternate distributions is still a subject of continuing investigation. In hydrology, double expenential type 1 or Gumbel extreme value type 1 distribution is most often used. Similarly there are many plotting position formulae. The primary objective of the report is to determine the most appropriate plotting position formula for the Gumbel EV-1 distribution.

Keeping in view the objective of the study, various plotting position formulae are reviewed. The details of Gumbel EV-1 distribution are given in Appendix I.

2.0 REVIEW

In graphical approach of flood frequency analysis the data of annual peaks are plotted on various probability papers using unbiassed probability plotting positions. The purpose of this plotting is three fold (i) to fit the distribution, (ii) to identify the outliers and (iiii) to assess the goodness of fit. Various probability papers and plotting position formulae are reviewed in the subsequent sections.

2.1 Probability Papers

The primary purpose of plotting set of observed observations on probability paper is to simplify the inspection of their distribution. Probability paper for a given probability of exceedence $P(X \geq x)$ or probability of non exceedence $P(X \leq x)$ is so designed that a plot of P(X) against X is a straight line. If a set of data follows a straight line on probability paper then the data is said to follow this distribution. Thus the process of extraploting the probability plot to find the flood magnitude of desired higher return period is simplified, since straight line can be extended easily. The probability paper can be preapred for any distribution based only on two parameters of the distribution. Any additional parameter such as coefficient of skewness must be constant. The probability paper for Normal (/log normal) and Gumbel EV-1 distribution are commercially available. In the probability paper only probability of exceedence or non exceedance P(X) is given. This P(X) can be changed to return period or reduced variate through simple equations, explained later in the report.

As the Pearson type III distribution (log Pearson type III) has got the coefficient of skewness varying as a parameter, in addition to mean and standard deviation, its probability paper is not generally commercially available. It will require one paper for each value of skewness.

The probability papers for normal and Gumbel EV-1 distributions are given in Fig. 1 and 2 respectively.

2.2 Plotting Position

Determining the probability to assign a data point is commonly referred to as determining the plotting position. Probability plotting of hydrologic data requires that individual observations or data points be independent of each other and the sample data be representative of the population. Gumbel (1958) states the following five criteria for plotting position relationships

1. The plotting positions must be such that all sample members may be plotted.
2. The return period of a value equal to or larger than the largest observation and that of a value equal to or smaller than the smallest observation should converge towards N, the number of observations, and 1 respectively. He noted that this condition is not fulfilled in Hazen's method.

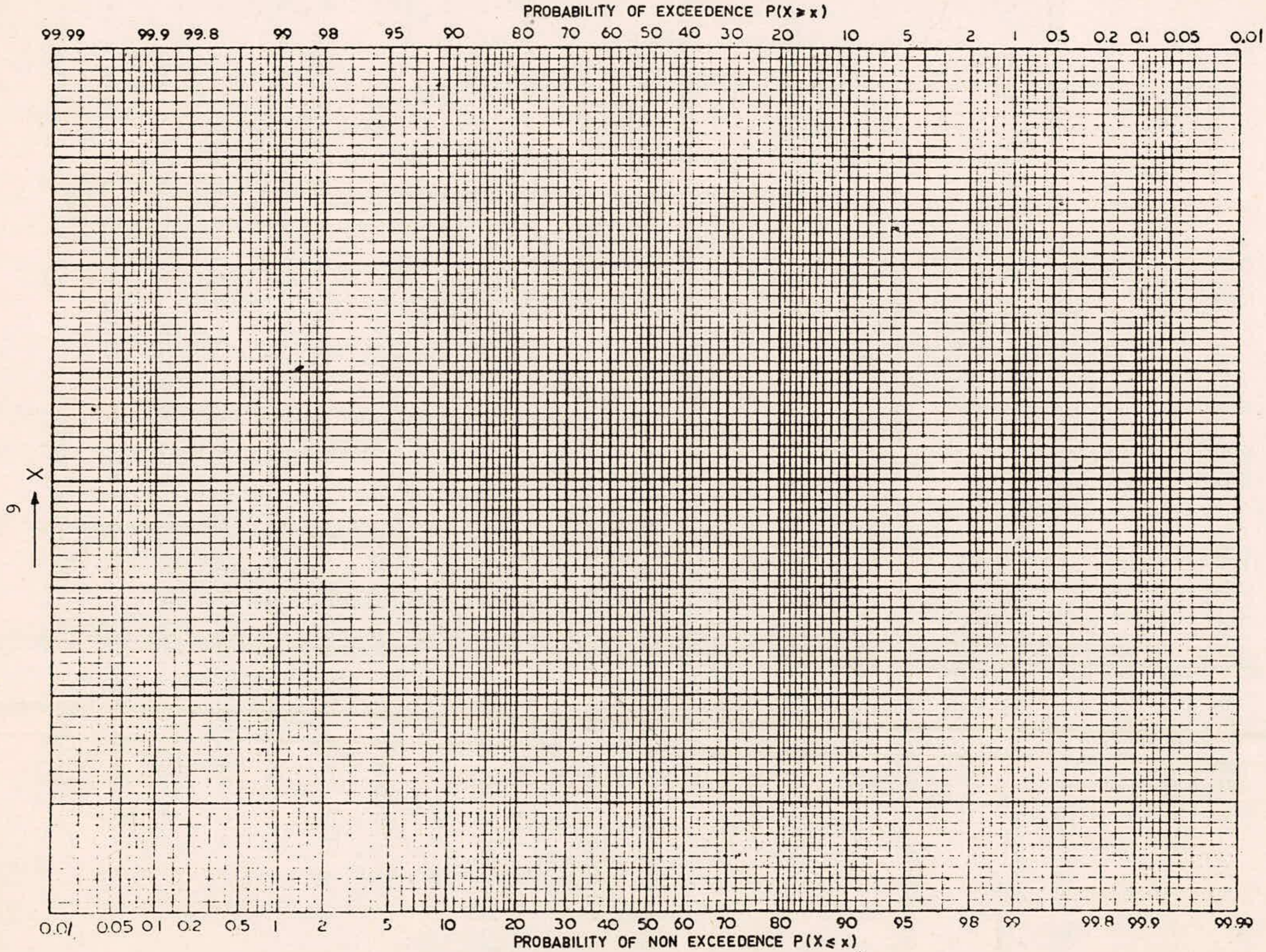

FIG. 1—NORMAL PROBABILITY PAPER

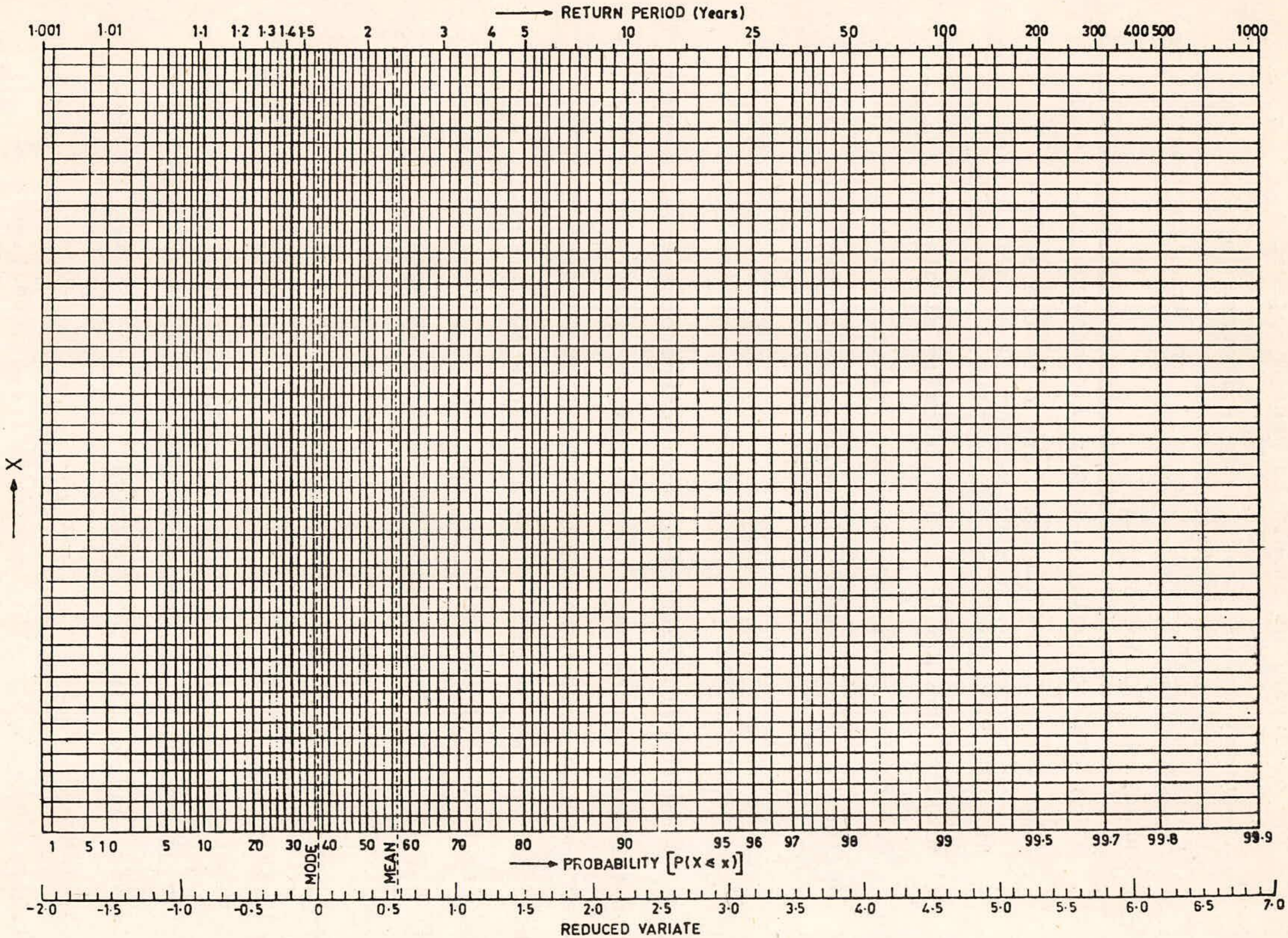

FIG. 2 — GUMBEL EXTREME VALUE TYPE I DISTRIBUTION

3. The observations should be equally placed on frequency scale, that is the difference between the plotting positions of the (i+1) th and ith observation should be idenpendent of i.

4. The plotting position should lie between (i+1)/N and i/N, where i is the rank and N is the sample size.

5. The plotting position should have an intuitive meaning and should be analytically simple.

One of the most commonly used plotting position is due to Weilbull. The form of the Weibull's plotting position is given as

$$P(X) = \frac{i}{N+1} \qquad \ldots (3)$$

where P (X) is the probability of exceedence of a given event, X, when the data are ranked from the larged (i=1) to the smallest i = N in descending order. The Wibull plotting position formula meets all the requirements specified by Gumbel as (i) all of the observations can be plotted since the plotting positions range from 1/(N+1) which is greater than zero to N/(N+1) which is less than 1, (ii) the return period of largest value is (N+1)/1 which approaches N as N gets large and the return period of smallest value is (N+1)/N which approaches 1 as N gets large, (iii) the difference between the plotting posicion of the (i+1) ih and i th observation is 1/(N+1) for all values of i and N, (iv) the relationship i/(N+1) lies between the (i-1)/N and i/N for all values of i and N and (v) the simplicity of Weibull relationship fulfills condition 5.

Although Weibull's plotting position is widely used in our country and USA, its use is discouraged as it is considered to be a biassed plotting position formula which necessitates the critical study of above five conditions given by Gumbel.

Condition 1: No exception can be taken to this condition, and in fact it is necessary

Condition 2: This condition is not in keeping with the statistical fact. This is most unleading condition and also appears to have played a major part in the adoption of Weibull formula (Cunnane, 1978). It is not necessary that the highest observation in a sample should converge towards a return period of N years. This may be true if the length of sample is very vary large, but in case of small samples ($<$ 30 years), which is generally the case, it is not true.

Condition 3: This condition is neither necessary nor desirable.

Condition 4: This condition is also not necessary.

Condition 5: This condition although desirable, is not reconcilable with any mathematical derivation as simplicity can not be used in the same way, as can, for instance, a boundary or initial condition and consequently can play no part in the rationale development of a formula. Numerous methods have been proposed for determination of plotting position. Most of them are empirical. Various plotting position formulae used in frequency analysis are described below .

2.2.1 California plotting position

It is the earliest (1923) formula for computing plotting positions. Use of this formula is known as the California method since it was first employed to plot flow data of California streams.

The probability of exceedence $P(X \geqslant x)$ is given by

$$P(X \geqslant x) = i/N \qquad \ldots. (4)$$

where i is the rank, if the data are arranged in descending order and N is the total number of data points.

Chow (1953) demonstrated theoretically that this method is suitable for plotting annual exceedence series or partial duration series. However this formula plots data at the edges of group intervals and produces a probability of 100 percent which can not be plotted on a probability paper.

2.2.2 Hazen plotting position

Since California formula plots data at the edges of group intervals and produces a probability of 100 percent which can not be plotted on a probability paper, it was gradually replaced by Hazen formula which plots data at the centres of the group intervals. The probability of exceedence

$P(X \geq x)$ is given by

$$P(X \geq x) = \frac{2\,i - 1}{2\,N}$$

$$= \frac{i - 0.5}{N} \qquad \ldots (5)$$

2.2.3 Weibull plotting position

The most practical plotting position which fully satified all five of Gumbel's conditions is Weibull plotting position. The probability of exceedence $P(X \geq x)$ is gven by

$$P\ (X \geq x) = \frac{i}{N + 1} \qquad \ldots (6)$$

The property of equation (6) is that for $i = 1$ it gives $P(X_1) = 1/(N+1)$ and for $i = N$ it yields $P(X_N) = N/(N+1)$, the first value being very close to the frequency $1/N$ and the last value being smaller than one. The use of equation (6) dominates at present the hydrologic applications.

The inverse of $P(X\ \ x)$ is called the return period or the recurrence interval which is given by

$$R\ (X) = \frac{1}{P(X \geq x)} \qquad \ldots (7)$$

The return period for the plotting position of equation (6) gives for $i = 1$ and $i = N$ the values of $R\ (X_1) = N + 1$ and $R\ (X_N) = 1 + (1/N)$. For the sample size of $N=100$ the return periods are 101 and 1.01 respectively. The first value should be 100, but the difference is negligible, except for very small samples, and the second value is greater than one but very closase to it which is what it should be.

2.2.4 Beard plotting position

The probability of exceedence of P(X x) is given by

$$P(X \geq x) = (i-0.31)/(N+0.38) \quad \ldots (8)$$

2.2.5 Chagodajew plotting position

This is an empirical formula, commonly used in the U S S R for Pearson type III distribution. The probability of exceedence $P(X \geq x)$ is given by

$$P(X \geq x) = (i - 0.3)/(N + 0.4) \quad \ldots (9)$$

Chegodajew formula is mathematical approximation of Beard formula.

2.2.6 Gringorton plotting position

In order to simplify the visual inspection of a plotted set of ordered observations on extremal probability paper, Gringorton (1963) recommended following formula for computing plotting positions

$$P(X \geq x) = (i - 0.44)/(N+0.12) \quad \ldots (10)$$

2.2.7 Blom plotting position

The probability of exceedence $P(X \geq x)$ is given by

$$P(X \geq x) = (i - 3/8)/(N + 1/4) \quad \ldots (11)$$

The Blom's plotting position formula has been proved to be unbiassed i.e. the average value of an ordered event, say X_i, considered over many number of samples of the same size would lie on the population line when plotted against the Blom's plotting position value on normal probability paper.

2.2.8 Tukey plotting position

The probability of exceedence $P(X \geqslant x)$ is given by

$$P(X \geqslant x) = (i-0.333)/N+0.333) \quad \ldots (12)$$

2.2.9 Benard plotting position

The probability of exceedence $P(X \geqslant x)$ is given by

$$P(X \geqslant x) = (i-0.3)/(N+0.2) \quad \ldots (13)$$

2.2.10 Foster plotting position

The probability of exceedence $P(X \geqslant x)$ is given by

$$P(X \geqslant x) = (i-0.5)/N \quad \ldots (14)$$

In fact, this is identical to Hazen plotting posotic formula.

2.2.11 Cunnane plotting position

The probability of exceedence $P(X \geqslant x)$ is given by

$$P(X \geqslant x) = (i-0.4)/(N+0.2) \quad \ldots (15)$$

2.2.12 Adamowski plotting position

The probability of exceedence $P(X \geqslant x)$ is given by

$$P(X \geqslant x) = (i-0.25)/(N+0.5) \quad \ldots (16)$$

All of the above plotting position formulae are special cases of general formula

$$P(X \geqslant x) = (i-\alpha)/(N+1-2\alpha) \quad \ldots (17)$$

where α varies from 0 to 1.

These formula can also be expressed as a special case of the following expression

$$P(X \geqslant x) = (i-a)/(N+b) \quad \ldots (18)$$

where a and b are constants.

2.3 Summary of Different Plotting Formulae

Different plotting position formulae have been summarized by Adamowiski (1981). The summary has been reproduced in Table 1. From the table it can be seen that for sample size 10, the return period for the highest value varies from 10 years (California) to 20 years (Hazen). This shows the implication involved, if the proper plotting position formula is not used.

TABLE - 1

VARIOUS PLOTTING POSITION FORMULAE

S.No.	Plotting position formula	Year	$P(X \geqslant x)$	Values of a and b in the formula $P(X \geqslant x) = \frac{(i - a)}{(N + b)}$	Value of α in the formula $P(X \geqslant x) = \frac{i - \alpha}{N+1-2}$	Value of $P(X \geqslant x)$ and T for i=1 and N=10 $P(X \geqslant x)$	$T = \frac{1}{P(X \geqslant x)}$
1.	California	1923	i/N	a=0.0, b=0.0	-	0.1	10
2.	Hazen	1930	(i-0.5)/N	a=0.5, b=0.0	0.5	0.05	20.0
3.	Weibull	1939	i/(N+1)	a=0.0, b=1.0	0.0	0.091	11.0
4.	Beard	1953	(i-0.31)/ (N+0.38)	a=0.31, b=0.38	0.31	0.066	15.0
5.	Chegodajew	1955	(i-0.3)/ (N+0.4)	a=0.3, b=0.4	0.3	0.067	14.9
6.	Gringorton	1963	(i-0.44)/ (N+0.12)	a=0.44, b=0.12	0.44	0.055	18.1
7.	Blom	1958	(i-3/8)/ (N+1/4)	a=3/8, b=1/4	3/8	0.0609	16.4
8.	Tukey	1962	(i-1/3)/ (N+1/3)	a=1/3, b=1/3	1/3	0.0648	15.4
9.	Benard	1953	(i-0.3)/ (N+0.2)	a=0.3, b=0.2	-	0.068	14.7
10.	Foster	1936	(i-0.5)/N	a=0.5, b=0	0.05	0.05	20
11.	Cunnane	1978	(i-0.4)/ (N+0.2)	a=0.4, b=0.2	0.4	0.053	17.2
12.	Adamowski	1981	(i-0.25)/ (N+0.5)	a=0.25, b=0.5	0.25	0.071	14.0

3.0 PROBLEM DEFINITION

In flood frequency analysis it is required to estimate the values of probabilities based on plotting formula. All of the many existing formula provide different results, particularly at the tails of the distribution. The existing practice in selection of a particular formula is arbitrary and often weibull's formula is recommended which provides biassed and conservative results (Adamowski, 1981).

Gringorton plotting position formula has been recommended by Gringerton (1963) and Cunnane (1978) for Gumbel EV-1 distribution on theoretical basis. Adamowski (1981) advocates for Adamowski plotting position formula for Gumbel EV-1 distribution for high values of probability of exceedence based on mean square error criteria. Similarly other plotting positions also have been recommended for one reason or the other. The main objective of this study is to compare various plotting position formulae on the basis of some statistical criteria, using synthetically generated Gumbel EV-1 distributed random numbers.

4.0 PROPOSED METHODOLOGY

In order to compare different plotting position formulae, the reduced variates given by a particular plotting position formula are compared with the expected value of reduced variates derived from the generated data. The comparison is made on the basis of:

(i) Over all fit

Total sum of squares of error has been used as overall fit performance criterion.

(ii) Fit in the upper tail region

The following three criteria have been used to judge the performance of fit in the upper tail region

a) Absolute error in the highest reduced variate,

b) Sum of squares of error in top 3 reduced variates and

c) Sum of squares of error in top 6 reduced variates

(iii) Fit in the lower tail region

The following three criteria have been used to judge the performance in the lower tail region.

a) Absolute error in the lowest reduced variate,

b) Sum of squares of error in bottom 3 reduced variates and

c) Sum of squares of error in bottom 6 reduced variates.

The study has been carried out for different length of samples. The sample lengths have been taken as 10, 20, 30 40, 50, 60, 70, 80, 90 and 100. Various steps, involved are explained in detail as below:

1. Generate 50,000 Gumbel extreme value type 1 (EV 1)

distributed random numbers. 50,000 has been chosen in order to have sufficient samples of different lengths. The location parameter u, and scale parameter α have been taken as 0.0 and 1.0 respectively. u and being 0.0 and 1.0 respectively, the EV 1 distributed random numbers are directly the reduced variates. The algorithm used for the generation is as follows:

a) Generate 50,000 uniformly distributed random numbers between 0.0 and 1.0.

b) Transform uniformly distributed random numbers to EV 1 distributed random numbers by the following formula

$$X = u + \alpha(-\ln(-\ln(Z)) \quad \ldots (19)$$

where,

$u = 0.0$

$\alpha = 1.0$

X = EV 1 distributed random numbers

Z = Uniformly di tributed random numbers

2. Get samples of different lengths (10, 20, 30, 40, 50, 60, 70, 80, 90 and 100) from the generated EV 1 distributed reduced variates e.g. if the sample length is 10 then one will be getting 50,000/10 5,000 samples.

3. Arrange the samples in descending order

4. Get the expected value of reduced variates for different sample sizes

$$E(Y_i) = \frac{\sum_{j=1}^{N} Y_{i,j}}{N} \qquad \ldots (20)$$

where,

$E(Y_i)$ = Expected valaue of ith reduced variate,

$Y_{i,j}$ = i th reduced variate for the jth sample and

N = No. of samples

The number of reduced variates will be equal to the size of the sample. $E(Y_i)$ can be considered as unbiassed probability plotting position.

5. Calculate probabilities of non exceedence from Weibull plotting position formula

$$P_i = 1 - \frac{i}{N + 1} \qquad \ldots (21)$$

where,

P_i = Probability of non exceedence for the ith variable,

i = Rank, if the series is arranged in descending order

N = Sample size

6. Calculate reduced variates corresponding to the probabilities of non exceedence.

$$Y_i = - \ln (-\ln(P_i))$$

where, Y_i is the ith reduced variate, and P_i is the ith probability of non exceedence.

7. Calculate (i) total sum of squares of error, (ii) absolute error in highest reduced variate, (iii) sum of squares of error in top 3 reduced variates, (iv) sum of squares of error in top 6 reduced variates,

(v) absolute error in lowest reduced variate,

(vi) sum of squares of error in bottom 3 reduced variates, and (vii) sum of squares of error in bottom 6 reduced variates.

The following equations are used

$$SSE = \sum_{i=1}^{N} (E(Y_i) - Y_i)^2 \qquad \ldots (22)$$

$$ABSY1 = ABS \, | E(Y_1) \quad Y_1 | \qquad \ldots (23)$$

$$SSET3 = \sum_{i=1}^{3} (E(Y_i) - Y_i)^2 \qquad \ldots (24)$$

$$SSET6 = \sum_{i=1}^{6} (E(Y_i) - Y_i)^2 \qquad \ldots (25)$$

$$ABSYN = ABS \, | E(Y_N) - Y_N | \qquad \ldots (26)$$

$$SSEB3 = \sum_{i=N-2}^{N} (E(Y_i) - Y_i)^2 \qquad \ldots (27)$$

$$SSEB6 = \sum_{i=N-5}^{N} (E(Y_i) - Y_i)^2 \qquad \ldots (28)$$

where,

SSE = Total sum of squares of error,

ABSY1 = Absolute error in highest reduced variate,

SSET3 = Sum of squares of error in top 3 reduced variates,

SSET6 = Sum of squares of error in top 6 reduced variates,

SSEB3 = Sum of squares of error in bottom 3 reduced varaites,

SSEB6 = Sum of squares of error in bottom 6 reduced variates

ABSYN = Absolute error in lowest reduced variates.

$E(Y_i)$ = Expected value of ith reduced variate ,

Y_i = ith reduced variate from the plotting position formula and

N = Sample size

8. Repeat steps from 5 to 7 for other eleven plotting positions formulae.

5.0 ANALYSIS AND RESULTS

The suitability of following plotting position formulae for use of EV 1 in flood frequency analysis, has been judged using seven different criteria as stated earlier.

i) Weibull, ii) Beard, iii) Hazen, iv) Benard, v) Blom, vi) Tukey, vii) Gringorton viii) California, ix) Chegodajew x) Foster, xi) Cunnane and xii) Adamowski

The sample sizes have been taken as 10, 20, 30, 40, 50, 60, 70, 80 , 90 and 100 from the generated 50,000 EV 1 distributed randum numbers.

The results have been tabulated in Table 2 to 8.

The expected value of reduced variates $E(Y_i)$ and reduced variates for different plotting positions have been plotted in fig.3 to 5 for sample size 10 and in Fig. 6 to 8 for sample size 30. For other sample sizes the graphs have not been plotted.

5.1 Overall Fit

For all the sample sizes, Gringorton plotting position formula gives minimum sum of squares of error.

The root mean square error, defined, as, square root of mean of total sum of squares of error, keeps on reducing as the sample size increases. The values of root mean square error for different sample sizes in the case of Gringorton plotting position formula are tabulated below.

TABLE - 2

COMPARISON OF DIFFERENT PLOTTING POSITION FORMULAE ON THE BASIS OF TOTAL SUM OF SQUARES OF ERROR

Plotting position formula	Sample Size									
	10	20	30	40	50	60	70	80	90	100
Weibull	0.3847	0.4486	0.4865	0.4941	0.4963	0.5177	0.5049	0.4967	0.5047	0.5011
Beard	0.0662	0.0701	0.0766	0.0752	.0.0737	0.0300	0.0747	0.0707	0.0734	0.0715
Hazen	0.0263	0.0209	0.0171	0.0175	0.0197	0.0169	0.0212	0.0236	0.0227	0.0247
Benard	0.1207	0.1067	0.1067	0.1015	0.0975	0.1023	0.0962	0.0907	0.0933	0.0898
Blom	0.0307	0.0286	0.0307	0.0292	0.0284	0.0316	0.0291	0.0268	0.0284	0.0275
Tukey	0.0515	0.0529	0.0578	0.0563	0.0549	0.0602	0.0558	0.0524	0.0547	0.0530
Gringorton	0.0155	0.0102	0.0088	0.0080	0.0084	0.0085	0.0093	0.0092	0.0095	0.0099
California	2.3062	1.6948	1.4549	1.3054	1.2016	1.1410	1.0787	1.0221	0.9953	0.9531
Chegodajew	0.0731	0.0781	0.0354	0.0841	0.0825	0.0393	0.0836	0.0794	0.0323	0.0802
Foster	0.0263	0.0209	0.0171	0.0175	0.0197	0.0169	0.0212	0.0236	0.0227	0.0247
Cunnane	0.0221	0.0134	0.0191	0.0177	0.0173	0.0194	0.0130	0.0165	0.0177	0.0172
Adamowski	0.1122	0.1239	0.1352	0.1345	0.1329	0.1420	0.1346	0.1293	0.1330	0.1304

TABLE - 3

COMPARISON OF DIFFERENT PLOTTING POSITION FORMULAE ON THE BASIS OF ABSOLUTE ERROR IN HIGHEST REDUCED VARIATE

Plotting position formula	Sample size									
	10	20	30	40	50	60	70	80	90	100
Weibull	0.5395	0.5652	0.5851	0.5822	0.5717	0.5863	0.5726	0.5593	0.5730	0.5580
Beard	0.2134	0.2170	0.2294	0.2226	0.2098	0.2229	0.2082	0.1941	0.2017	0.1916
Hazen	0.0800	0.0908	0.0832	0.0923	0.1065	0.944	0.1099	0.1245	0.1119	0.1277
Benard	0.2464	0.2407	0.2499	0.2417	0.2279	0.2404	0.2252	0.2108	0.2235	0.2078
Blom	0.1242	0.1229	0.1337	0.1261	0.1129	0.1256	0.1106	0.0963	0.1092	0.0936
Tukey	0.1825	0.1844	0.1961	0.1891	0.1761	0.1891	0.1743	0.1601	0.1730	0.1575
Gringorton	0.0242	0.0180	0.0271	0.0188	0.0050	0.0174	0.0022	0.0123	0.0005	0.0152
California	0.6398	0.6153	0.6184	0.6072	0.5917	0.6030	0.5869	0.5718	0.5841	0.5680
Chegodajew	0.2266	0.2307	0.2432	0.2367	0.2239	0.2371	0.2223	0.2083	0.2213	0.2058
Foster	0.0800	0.0908	0.0832	0.0923	0.1065	0.0944	0.1099	0.1245	0.1119	0.1277
Cunnane	0.0871	0.0840	0.0941	0.0862	0.0728	0.0854	0.0703	0.0560	0.0688	0.0532
Adamowski	0.2379	0.2958	0.3098	0.3038	0.2914	0.3048	0.2903	0.2763	0.2894	0.2741

TABLE - 4

COMPARISON OF DIFFERENT PLOTTING POSITION FORMULAE ON THE BASIS OF SUM OF SQUARES OF ERROR IN TOP 3 REDUCED VARIATES

Plotting position formula	Sample Size									
	10	20	30	40	50	60	70	80	90	100
Weibull	0.3588	0.4081	0.4330	0.4355	0.4312	0.4483	0.4286	0.4228	0.4224	0.4217
Beard.	0.0612	0.0617	0.0675	0.0655	0.0624	0.0677	0.0592	0.0570	0.0560	0.0558
Hazen	0.0073	0.0083	0.0070	0.0086	0.0114	0.0090	0.0122	0.0155	0.0137	0.0166
Benard	0.0864	0.0779	0.0814	0.0779	0.0737	0.0790	0.0694	0.0657	0.0655	0.0650
Blom	0.0241	0.0218	0.0243	0.0229	0.0212	0.0239	0.0189	0.0182	0.0171	0.0176
Tukey	0.9462	0.0455	0.0500	0.0481	0.0454	0.0498	0.0425	0.0408	0.0398	0.0398
Gringorton	0.0042	0.0020	0.0021	0.0019	0.0023	0.0023	0.0013	0.0025	0.0011	0.0026
California	0.5730	0.5090	0.5038	0.4884	0.4734	0.4843	0.4582	0.4483	0.4451	0.4425
Chegodajew	0.0681	0.0693	0.0756	0.0736	0.0703	0.0760	0.0670	0.0646	0.0637	0.0634
Foster	0.0073	0.0083	0.0070	0.0086	0.0114	0.0090	0.0122	0.0155	0.0137	0.0166
Cunnane	0.0141	0.0114	0.0129	0.0119	0.0109	0.0126	0.0091	0.0091	0.0078	0.0087
Adamowski	0.1064	0.1119	0.1213	0.1193	0.1153	0.0123	0.1116	0.1083	0.1075	0.1068

TABLE - 5

COMPARISON OF DIFFERENT PLOTTING POSITION FORMULAE ON THE BASIS OF SUM OF SQUARES OF ERROR IN TOP 6 REDUCED VARIATES

Plotting position formula	Sample size									
	10	20	30	40	50	60	70	80	90	100
Weibull	0.3666	0.4223	0.4559	0.4620	0.4631	0.4820	0.4688	0.4592	0.4630	0.4572
Beard	0.0654	0.0659	0.0715	0.0700	0.0687	0.0742	0.0685	0.0643	0.0651	0.0624
Hazen	0.0101	0.0086	0.0071	0.0086	0.0117	0.0092	0.0130	0.0159	0.0145	0.0167
Benard	0.0988	0.0854	0.0875	0.0842	0.0818	0.0870	0.0804	0.0754	0.0759	0.0728
Blom	0.0277	0.0242	0.0263	0.0250	0.0245	0.0273	0.0243	0.0221	0.0222	0.0228
Tukey	0.0502	0.0489	0.0531	0.0517	0.0505	0.0551	0. 503	0.0467	0.0573	0.0451
Gringorton	0.0073	0.0031	0.0027	0.0025	0.0035	0.0036	0.0038	0.0040	0.0034	0.0037
California	0.6779	0.5638	0.5486	0.5314	0.5193	0.5292	0.5091	0.4940	0.4939	0.4949
Chegodajew	0.0724	0.0738	0.0800	0.0786	0.0771	0.0830	0.0770	0.0725	0.0734	0.0705
Foster	0.0101	0.0086	0.0071	0.0086	0.0117	0.0092	0.0130	0.0159	0.0145	0.0167
Cunnane	0.0175	0.0132	0.0143	0.0133	0.0133	0.0151	0.0133	0.0119	0.0118	0.0110
Adamowski	0. 112	0.1182	0.1278	0.1286	0.1251	0.1331	0.1253	0.1196	0.1210	0.1173

TABLE - 6

COMPARISON OF DIFFERENT PLOTTING POSITION FORMULAE ON THE BASIS OF ABSOLUTE ERROR IN LOWEST REDUCED VARIATE

Plotting position formula	sample size									
	10	20	30	40	50	60	70	80	90	100
Weibull	0.1205	0.1175	0.1107	0.1029	0.0978	0.0934	0.895	0.0864	0.0824	0.0836
Beard	0.0022	0.0113	0.0135	0.0114	0.0105	0.0143	0.0079	0.0069	0.0048	0.0075
Hazen	0.1021	0.0745	0.0651	0.0627	0.0603	0.0540	0.0583	0.0576	0.0584	0.0544
Benard	0.1087	0.0772	0.0668	0.0638	0.0611	0.0547	0.0589	0.0581	0.0588	0.0548
Blom	0.0336	0.0157	0.0112	0.0119	0.0118	0.0071	0.0129	0.0133	0.0150	0.0119
Tukey	0.0132	0.0019	0.0048	0.0032	0.0027	0.0068	0.0006	0.0001	0.0022	0.0007
Gringorton	0.0677	0.0449	0.0381	0.0372	0.0359	0.0304	0.0355	0.0354	0.0366	0.0330
California	1.2252	0.9895	0.8759	0.8055	0.7534	0.7083	0.6810	0.6536	0.6314	0.6074
Chegodajew	0.0024	0.0153	0.0171	0.0148	0.0137	0.0174	0.0110	0.0099	0.0077	0.0104
Foster	0.1021	0.0745	0.0651	0.0627	0.0603	0.0540	0.0583	0.0576	0.0584	0.0544
Cunnane	0.0463	0.0266	0.0213	0.0213	0.0208	0.0159	0.0213	0.0216	0.0231	0.0198
Adamowski	0.0247	0.0345	0.0347	0.0313	0.0295	0.0326	0.0257	0.0243	0.0217	0.0242

TABLE - 7

COMPARISON OF DIFFERENT PLOTTING POSITION FORMULAE ON THE BASIS OF SUM OF SQUARES OF ERROR IN BOTTOM 3 REDUCED VARIATES

Plotting position formula	Sample Size									
	10	20	30	40	50	60	70	80	90	100
Weibull	0.0181	0.0194	0.0177	0.0159	0.0146	0.0139	0.0121	0.0116	0.0107	0.0109
Beard	0.0003	0.0002	0.0003	0.0003	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002
Hazen	0.0149	0.0074	0.0054	0.0048	0.0043	0.0037	0.0041	0.0039	0.0039	0.0034
Benard	0.0189	0.0033	0.0059	0.0050	0.0045	0.0039	0.0042	0.0040	0.0040	0.0035
Blom	0.0023	0.0004	0.0002	0.0001	0.0001	0.0001	0.0002	0.0002	0.0002	0.0001
Tukey	0.0007	0.0000	0.0000	0.0001	0.0001	0.0001	0.0000	0.0001	0.0001	0.0001
Gringorton	0.0071	0.0023	0.0019	0.0016	0.0015	0.0012	0.0015	0.0014	0.0015	0.0012
California	1.5983	1.0275	0.8024	0.6773	0.5922	0.5249	0.4850	0.4465	0.4170	0.3858
Chegodajew	0.0003	0.0003	0.0004	0.0004	0.0004	0.0004	0.0003	0.0003	0.0002	0.0003
Foster	0.0149	0.0074	0.0054	0.0048	0.0043	0.0037	0.0041	0.0039	0.0039	0.0034
Cunnane	0.0038	0.0010	0.0006	0.0005	0.0005	0.0003	0.0005	0.0005	0.0006	0.0004
Adamowski	0.0007	0.0016	0.0017	0.0016	0.0015	0.0015	0.0011	0.0011	0.1075	0.0011

TABLE - 8

COMPARISON OF DIFFERENT PLOTTING POSITION FORMULAE ON THE BASIS OF SUM OF SQUARES OF ERROR IN BOTTOM 6 REDUCED VARIATES

Plotting Position formula	Sample Size									
	10	20	30	40	50	60	70	80	90	100
Weibull	0.00205	0.0204	0.0192	0.0180	0.0166	0.0160	0.0139	0.0137	0.0124	0.0126
Beard	0.0030	0.0003	0.0003	0.0003	0.0003	0.0004	0.0002	0.0003	0.0002	0.0003
Hazen	0.0184	0.0084	0.0064	0.0052	0.0047	0.0040	0.0044	0.0040	0.0042	0.0036
Benard	0.0295	0.0109	0.0073	0.0057	0.0050	0.0042	0.0046	0.0042	0.0043	0.0037
Bloom	0.0051	0.0008	0.0004	0.0002	0.0002	0.0001	0.0002	0.0002	0.0002	0.0002
Tukey	0.0035	0.0003	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
Gringorton	0.0103	0.0037	0.0024	0.0018	0.0016	0.0013	0.0016	0.0014	0.0016	0.0013
California	1.6900	1.0576	0.8216	0.7905	0.6033	0.5342	0.4935	0.4536	0.4242	0.3925
Chegodajew	0.0029	0.0004	0.0005	0.0005	0.0005	0.0005	0.0003	0.0004	0.0003	0.0004
Foster	0.0184	0.0089	0.0064	0.0052	0.0047	0.0040	0.0044	0.0040	0.0042	0.0036
Cunnane	0.0067	0.0016	0.0009	0.0006	0.0005	0.0003	0.0005	0.0005	0.0006	0.0004
Adamowski	0.0032	0.0016	0.0018	0.0018	0.0017	0.0017	0.0013	0.0014	0.0011	0.0013

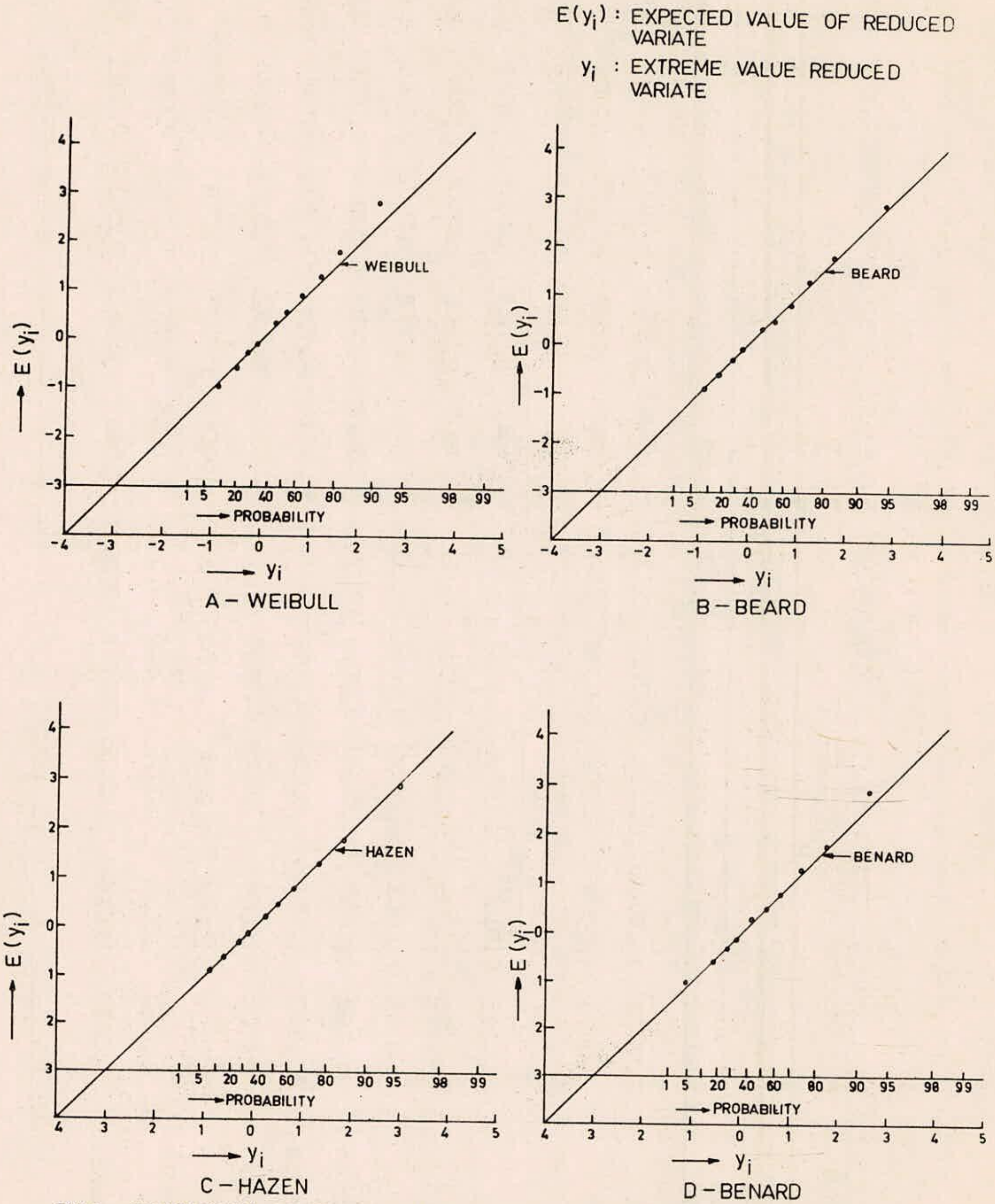

FIG. 3 — COMPARISON OF PROBABILITY PLOTS FOR DIFFERENT PLOTTING FORMULAE FOR SAMPLE SIZE 10

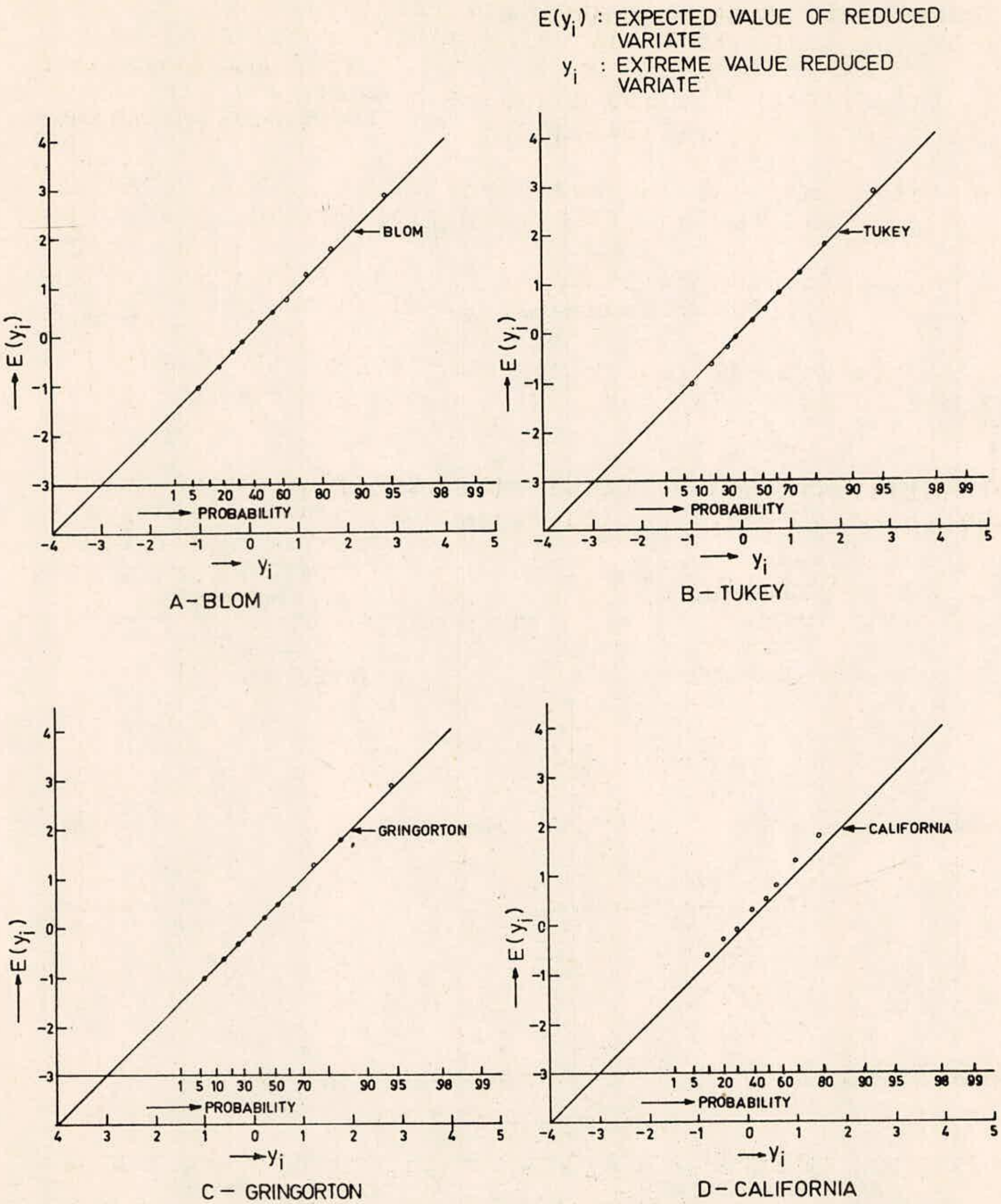

FIG. 4 — COMPARISON OF PROBABILITY PLOTS FOR DIFFERENT PLOTTING FORMULAE FOR SAMPLE SIZE 10.

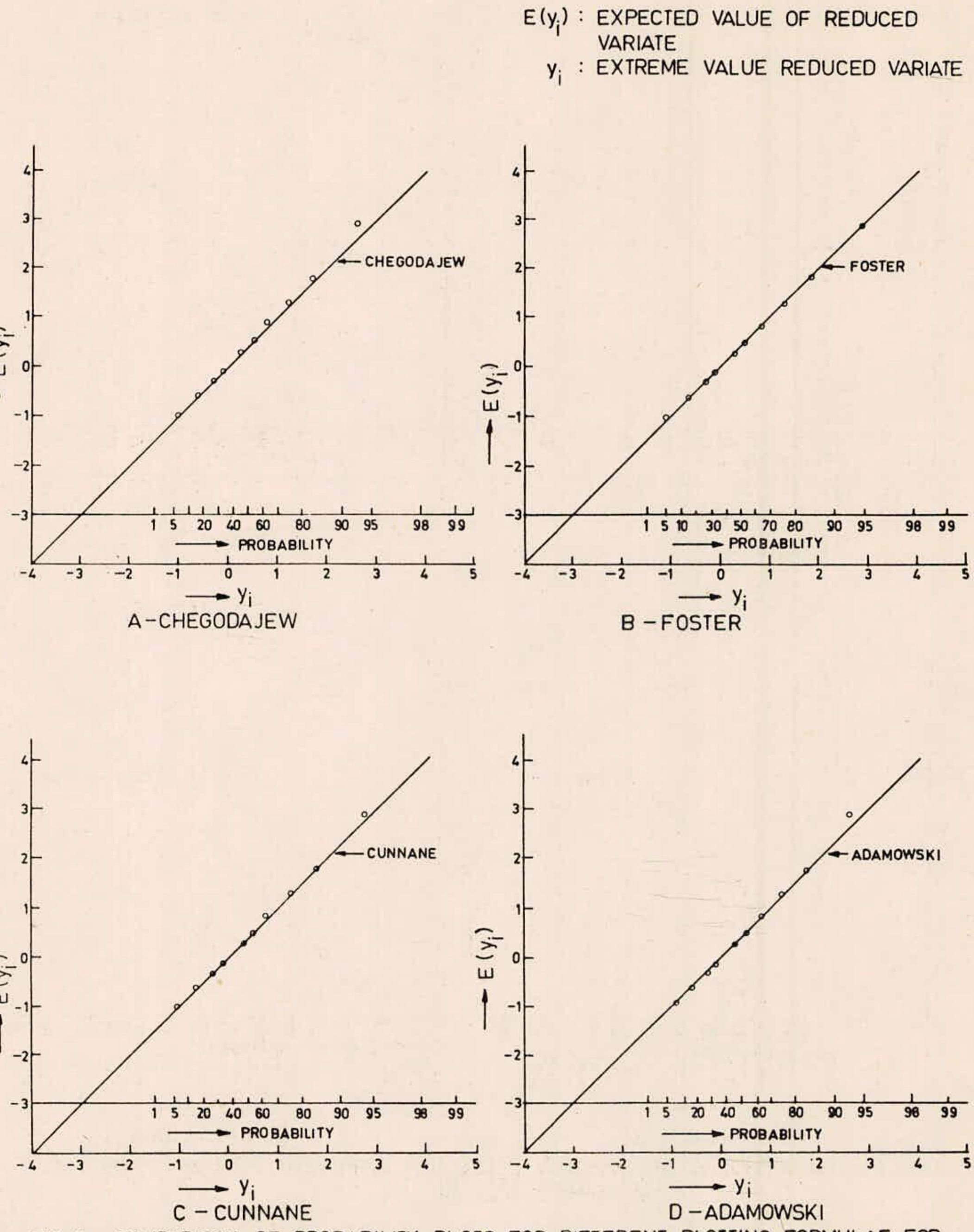

FIG. 5 — COMPARISON OF PROBABILITY PLOTS FOR DIFFERENT PLOTTING FORMULAE FOR SAMPLE SIZE 10

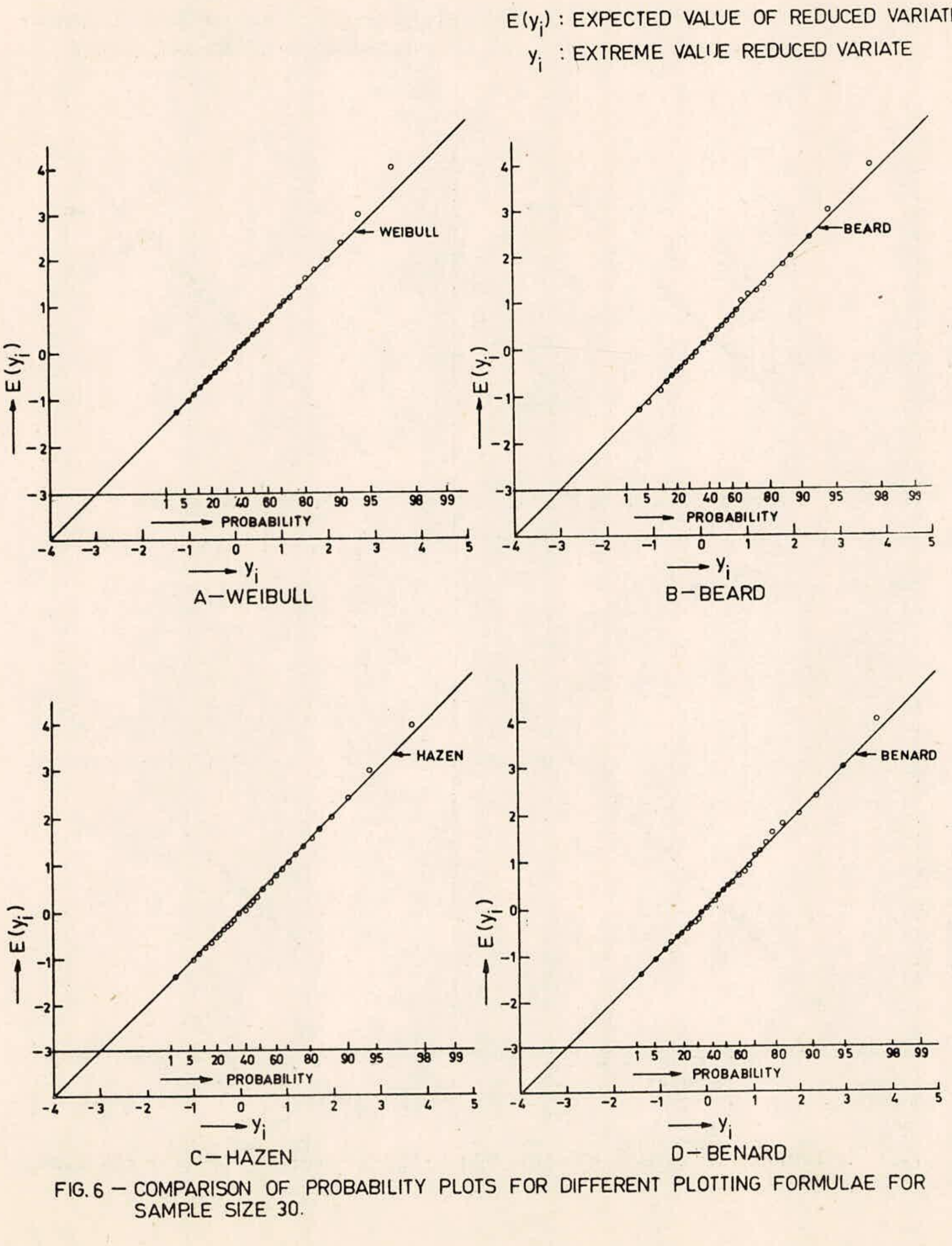

FIG. 6 — COMPARISON OF PROBABILITY PLOTS FOR DIFFERENT PLOTTING FORMULAE FOR SAMPLE SIZE 30.

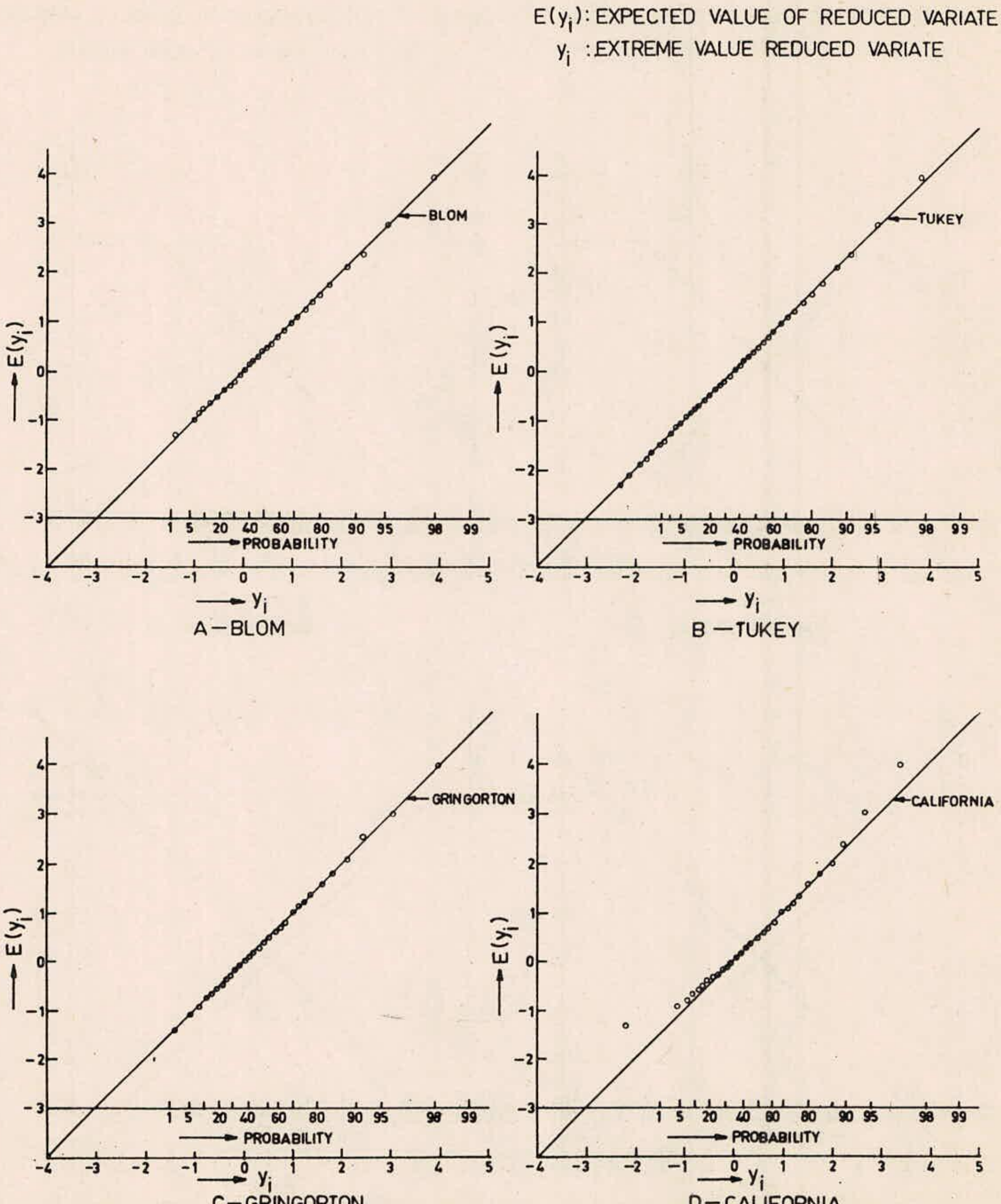

FIG. 7—COMPARISON OF PROBABILITY PLOTS FOR DIFFERENT PLOTTING FORMULAE FOR SAMPLE SIZE 30.

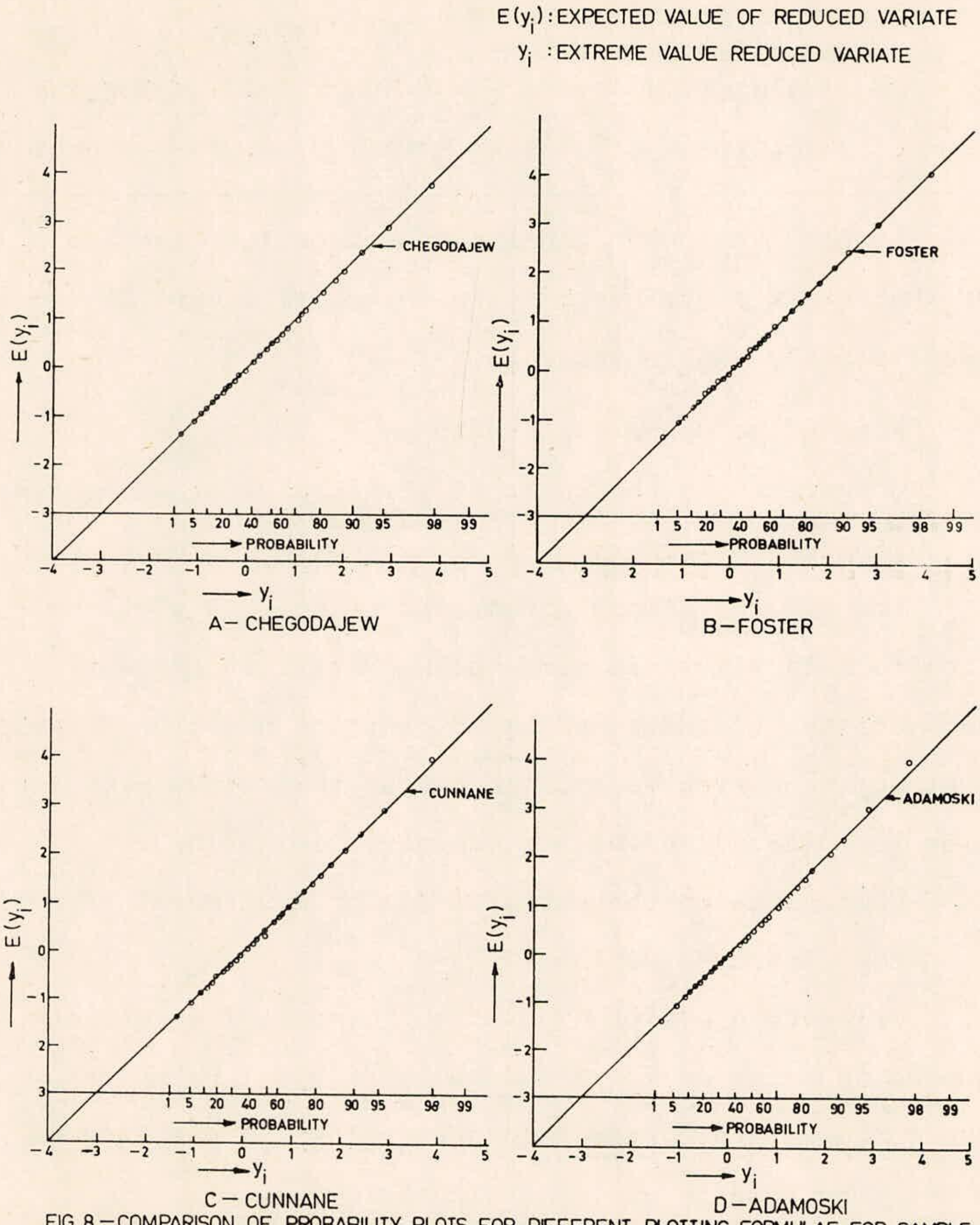

FIG. 8 – COMPARISON OF PROBABILITY PLOTS FOR DIFFERENT PLOTTING FORMULAE FOR SAMPLE SIZE 30

Sample size	Root mean square error	Sample size	Root mean square error
10	0.0394	60	0.0119
20	0.0226	70	0.0115
30	0.0171	80	0.0107
40	0.0141	90	0.0103
50	0.0120	100	0.0099

Cunnane and Hazen plotting position formulae are better than other plotting position formulae, except Gringorton. The results are given in Table 2.

5.2 Fit in the Upper Tail Region

5.2.1 Comparison on the basis of absolute error in highest reduced variate

For all the sample sizes, the Gringorton plotting position formula gives minimum absolute error in highest reduced variate. Cunnane and Hazen plotting position formulae give, almost same results and are better than other plotting position formulae. The results are given in table 3.

5.2.2 Comparison on the basis of sum of squares of error in top 3 reduced variates

Gringorton plotting position formula gives minimum sum of squares of error in top 3 reduced variates. Hazen and Cunnane formulae are better than other plotting position formulae. The results are given in Table 4.

5.2.3 Comparison on the basis of sum of squares of error in top 6 reduced variates

Gringorton plotting position formula is more close to the unbiassed plotting positions (expected value of reduced variates) in case of all the sample sizes. Hazen and Cunnane plotting position formulae are better than other plotting position formulae. The results are given in Table 5.

5.3 Fit in the Lower Tail Region

5.3.1 Comparison on the basis of absolute error in lowest reduced variate

Tukey plotting position formula gives minimum absolute error in the lowest reduced variate for all the sample sizes, except sample size 10. Beard plotting position formula is the second best in most of the sample sizes. The results are given in table 6.

5.3.2 Comparison on the basis of sum of squares of error in bottom 3 reduced variates

Tukey plotting position formula gives minimum sum of squares of error in bottom 3 reduced variates. Beard and Blom formulae are better than other plotting position formulae. The results are given in Table 7.

5.3.3 Comparison on the basis of sum of squares of error in bottom 6 reduced variates

Tukey plotting position formula is more close to the unbiassed plotting positions in case of all the sample sizes except sample size 10. Blom plotting position is the second best. The results are given in Table 8.

5.4 General Comments

Adamowski, (1981) concludes that 'when the flood frequency analysis is performed using Gumbel Type 1 distribution then the plotting formula developed in this paper provides good approximation to true exceedence probability at high values'.

This conclusion seems to be in error, because of the following reasons:

1. The conclusion has been arrived at only on the basis of sample size 10.
2. The values of probability of exceedence corresponding to expected value of reduced variates seem to be wrong. Instead of taking these values from Benjamin, et al (1970), these could be calculated directly from the following equation:

   $$G(Y) = 1 - e^{-e^{-Y}}$$

   where

   G(Y) is the probability of exceedence, and

   y is the reduced variate.

6.0 CONCLUSIONS

For Gumbel EV 1 distribution, various plotting position formulae have been compared with the unbiassed plotting positions (the expected value of reduced variates). Unbiassed plotting positions have been obtained from synthetically generated EV 1 reduced variates. The following conclusions can be drawn from the study.

1. The use of $E(Y_i)$, the expected value of the reduced variate order statistic, performs very well as plotting position and leads to an unbiased probability plot.
2. The plotting positions given by Gringorton formula, $P(X \geqslant x) = (i - 0.44)/(N + 0.12)$, are more close to the unbiased plotting positions for all the sample sizes (10-100) on the basis of (i) overall fit and (ii) fit in the upper tail region.
3. The performance of Cunnane and Hazen plotting position formulae is better than other formulae, except Gringorton, on the basis of above criteria.
4. The plotting positions given by Tukey formula, $P(X \geqslant x) = (i - 1/3)/(N + 1/3)$ are more close to the unbiased plotting positions in the lower tail region.

REFERENCES

1. Adamowski, K. (1981), 'Plotting Formula for Flood Frequency', Water Resources Bulletin, Vol. 17/2, pp. 197-202.

2. Beard, L.R. (1943), 'Statistical Analysis in Hydrology' Trans. Am. Soc. Civ. Engg., Vol. 108, pp. 1110-1160.

3. Benard, A. and Bos. Levenbach, E.C. (1953), 'The Plotting of Observations on Probability Paper', Statistica, Vol. 7, pp. 163 173.

4. Benjamin, J.R. and Cornell, C.A. (1970), 'Probability Statistics and Decision for Civil Engineers', McGraw Hill Book Co., New York, New York.

5. Blom, G. (1958), 'Statistical Estimates and Transformed Beta Variables' Wiley, New York, N.Y., pp. 68 75 and pp. 143 146.

6. California Department of Public Works (1923), 'Flow in California Streams', California Dept. Public Works, Div. Eng. Irrigation, Bulletin 5.

7. Chegodajew, (1955), 'Formulas for the Calculation of the Confidence of Hydrologic Quantities' by A.G. Alekseyev in 'Handbook of Applied Hydrology' edited by V T. Chow, 1984.

8. Chow, V.T. (1953), 'Frequency Analysis of Hydrological Data with Special Application to Rainfall Intensities', Univ. Ill.Eng. Exp. Stn. Bulletin, Ser. No.414, pp. 27-29.

9. Cunnane, C. (1978), 'Unbiassed Plotting Positions- A Review', Journal of Hydrology Vol. 37, pp. 205-222.

10. Foster, E.E. (1948), 'Rainfall and Runoff', The Macmillan Company, New York.

11 Gringorton, I.I. (1963), 'A Plotting Rule for Extreme Probabilitiy Paper', Journal of Geophys. Research, Vol. 68/3, pp 813-814.

12. Gumbel, E J. (1958), 'Statistics of Extremes', Columbia University Press, New York, N.Y., pp. 28 34.

13. Hazen, A. (1930),'Flood Flows - A study of Frequencies and Magnitudes', John Wiley and Sons Inc. New York.

14. Natural Environment Research Council (1975), 'Flood Studies Report', Vol. 1, pp. 81-84.

15. Tukey, J.W. (1962),'The Future of Data Analysis' Annals of Math. Statistics, Vol. 33/1, pp. 21-24.

16. Varshney, R.S. (1979), 'Engineering Hydrology', Nem Chand and Broths. , Roorkee, pp. 917.

17. Weibull W. (1939), A Statistical Theorey of Strength of Materials', Ing. Vet. AK. Handl., Stockholm, pp. 151.

18. Yevjevich, V. (1972), 'Probability and Statistics in Hydrology, Water Resources Publication, Fort Collins, Col., pp. 89-92.

APPENDIX - I

GUMBEL DISTRIBUTION

One of the most commonly used distributions in flood frequency analysis is the double exponential distribution (known as Gumbel distribution or extreme value type 1 or EV 1 distribution). The EV 1 distribution is defined as:

$$P(X \leq x) = \exp(-\exp(-(\frac{x-u}{\alpha}))) \qquad \ldots (1)$$

in which,

$P(X \leq x)$ = The probability of non exceedence,

u = Location parameter of the distribution, and

α = Scale parameter of the distribution

The above equation can be written in the reduced variate form as

$$P(X \leq x) = \exp(-\exp(-Y)) \qquad \ldots (2)$$

where,

$$Y = (X - u)/\alpha \qquad \ldots (3)$$

The reduced variate Y can be written in terms of return period, T, the inverse of $P(X \geq x)$ using equation (2) as:

$$Y = -\ln(-\ln(1 - \frac{1}{T})) \qquad \ldots (4)$$

$$= -\ln.\ln(\frac{T}{T-1})$$

or

$$X = u - \alpha.\ \ln.\ \ln(\frac{T}{T-1}) \qquad \ldots (5)$$

An alternative form of equation (5) is given as

$$X = \mu + K_T \cdot \sigma \qquad \ldots (6)$$

in which,

μ = The population mean of the annual maximum series

σ = The population standard deviation of the annual maximum series and

K_T = frequency factor

The frequency factor, K_T is defined as:

$$K_T = -\frac{\sqrt{6}}{\pi}\left(0.5772 + \ln.\ln\left(\frac{T}{T-1}\right)\right) \qquad \ldots (7)$$