Lecture Note on
Flood Probability (*Flood Frequency*)

Recalling that; $P = \frac{1}{T}$; where T =return period ,and P=the probability, using plotting position(p.p.) method;$P = \frac{m}{N+1}$; m=order number of events, and N=total number of events.

If the events were taken as discharges; a plot of Q vs. T yields a probability distribution. For small return periods, a simple best-fitting curve can e used to extrapolate the probability distribution . A log-scale for T is often preferable.

When larger T is involved, the frequency analysis of flood problems should be used to predict the extreme flood events.

The general equation of hydrologic frequency analysis is:

$$x_T = \bar{x} + k\,\sigma_x \quad \text{...(1)}$$

Where x_T=value of variate x of random hydrologic series with return period T; $\bar{x}$ =mean of the variate ; σ_x= standard deviation of variate; and k= frequency factor which depends on return period T; and the type of frequency distribution.

Some of the commonly used frequency functions:

1. **Gumbel's extreme-value distribution (widely used in UK).**
2. **Log-Pearson type III distribution (widely used in USA).**
3. **Log-Normal distribution .**

The following details is about the first type of distribution, which is the Extreme-Value type I .The method was firstly introduced by Gumbel (1941) and commonly known as Gumbel's distribution. The method was used to predict a flood peaks, maximum rainfall, maximum wind speed,….etc., and the annual series flows constitute a series of largest value of flows. أقصى تصريف خلال ٣٦٥ يوما وأن السلسلة الزمنية للتصاريف تشكل القيم القصوى لها

The probability of occurrence of an event equal or larger than value such x_o is:

$$P(x \geq x_o) = 1 - e^{-e^{y}} \quad \text{..(2)}$$

Y =a dimensionless variable , given by:

Y =α (x-a) ,in which; α =1.2825/ σ_x ,and a=$\bar{x}$ -0.45005 σ_x

Hence; y can be written as;

$$y = \frac{1.2825(x - \bar{x})}{\sigma_x} + 0.577 \quad \text{..................................(3)}$$

In practice, x is the value for given P that is required by Eq.(2) is transposed as;

$$y_P = -Ln\left[Ln(1-P)\right] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

And since $P = \frac{1}{T}$,so that;

$$y_T = -Ln\left[Ln\left(\frac{T}{T-1}\right)\right] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(5)$$

y_T =the value of y commonly called " reduced variate" العامل المختزل

Now ,rearranging Eq. (3) using x with return period T, and the reduced value of variate y to get the following:

$$x_T = \bar{x} + \frac{y_T - 0.577}{1.2825}\sigma_x$$

Thus; k= 0.7797 y_T - 0.45005

Or more suitable form;

$$k = \frac{y_T - 0.577}{1.2825} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(6)$$

These are the basic Gumbel's equations applicable to an infinite sample size(i.e. $N\rightarrow\infty$) .

Gumbel's Equations for Practical Use

Since practical annual data series of extreme events have finite lengths of records ;Eq.(6) must be modified by considering the finite N ,as given below;

$$x_T = \bar{x} + k\,\sigma_{n-1} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(7)$$

In which $\sigma_{n-1} = \sqrt{\frac{\sum(x-\bar{x})^2}{N-1}}$ and $\bar{x} = \frac{\sum x}{N}$

Therefore ; Eq.(6) became :

$$k = \frac{y_T - \overline{y_n}}{S_n} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(8)$$

$\overline{y_n}$ = reduced mean ; It was a function of sample size (N) and given in Table -1- ; for $N\rightarrow\infty$; $\overline{y_n}$ = 0.577

S_n= reduced standard deviation ; It was a function of sample size (N) and given in Table -2- ; for $N\rightarrow\infty$; S_n =1.2825.

The procedure used to estimate flood magnitude corresponding to given return period ,T are;

1. Collect discharge data and note the sample size N. Find $\bar{x}$ and σ_{n-1} .
2. Using Table -1- and Table -2- to determine $\overline{y_n}$ and S_n vs. N.
3. Find y_T for given T by Eq.(5).
4. Find k by Eq.(8).
5. Determine x_T using Eq.(7).

TABLE –1 - REDUCED MEAN $\bar{y}_n$ IN GUMBEL'S EXTREME VALUE DISTRIBUTION

N = sample size

N	0	1	2	3	4	5	6	7	8	9
10	0.4952	0.4996	0.5035	0.5070	0.5100	0.5128	0.5157	0.5181	0.5202	0.5220
20	0.5236	0.5252	0.5268	0.5283	0.5296	0.5309	0.5320	0.5332	0.5343	0.5353
30	0.5362	0.5371	0.5380	0.5388	0.5396	0.5402	0.5410	0.5418	0.5424	0.5430
40	0.5436	0.5442	0.5448	0.5453	0.5458	0.5463	0.5468	0.5473	0.5477	0.5481
50	0.5485	0.5489	0.5493	0.5497	0.5501	0.5504	0.5508	0.5511	0.5515	0.5518
60	0.5521	0.5524	0.5527	0.5530	0.5533	0.5535	0.5538	0.5540	0.5543	0.5545
70	0.5548	0.5550	0.5552	0.5555	0.5557	0.5559	0.5561	0.5563	0.5565	0.5567
80	0.5569	0.5570	0.5572	0.5574	0.5576	0.5578	0.5580	0.5581	0.5583	0.5585
90	0.5586	0.5587	0.5589	0.5591	0.5592	0.5593	0.5595	0.5596	0.5598	0.5599
100	0.5600									

TABLE –2 - REDUCED STANDARD DEVIATION S_n IN GUMBEL'S EXTREME VALUE DISTRIBUTION

N = sample size

N	0	1	2	3	4	5	6	7	8	9
10	0.9496	0.9676	0.9833	0.9971	1.0095	1.0206	1.0316	1.0411	1.0493	1.0565
20	1.0628	1.0696	1.0754	1.0811	1.0864	1.0915	1.0961	1.1004	1.1047	1.1086
30	1.1124	1.1159	1.1193	1.1226	1.1255	1.1285	1.1313	1.1339	1.1363	1.1388
40	1.1413	1.1436	1.1458	1.1480	1.1499	1.1519	1.1538	1.1557	1.1574	1.1590
50	1.1607	1.1623	1.1638	1.1658	1.1667	1.1681	1.1696	1.1708	1.1721	1.1734
60	1.1747	1.1759	1.1770	1.1782	1.1793	1.1803	1.1814	1.1824	1.1834	1.1844
70	1.1854	1.1863	1.1873	1.1881	1.1890	1.1898	1.1906	1.1915	1.1923	1.1930
80	1.1938	1.1945	1.1953	1.1959	1.1967	1.1973	1.1980	1.1987	1.1994	1.2001
90	1.2007	1.2013	1.2020	1.2026	1.2032	1.2038	1.2044	1.2049	1.2055	1.2060
100	1.2065									

How to verify whether the given data follow Gumbel's distribution or not……….. The following steps may be adopted:

1. Calculate x_T for some return periods $T < N$ using Gumbel's formula.(say 3 or 4 values of x)
2. Plot x_T vs. T on semi-log or log-log or Gumbel probability paper (T must plot on log-scale)..
3. The use of Gumbel probability paper results a straight line for x_T vs. T relationship.
4. The Gumbel's distribution has a property that the average annual series was located at T=2.33 years, when N is large. This value called "*Mean Annual Flood* " متوسط الفيضـان السنوي .In graphical plot this gives a mandatory point through which the line showing variation of x_T with T.
5. For given data , T may calculated from PP method ($\frac{N+1}{m}$) and for various recorded values , x of variate obtained from pp are plotted on graph descried above.
6. A good fit of observed data with theoretical variation line indicates applicability of Gumbel's distribution.

EXAMPLE 1 — *Annual maximum recorded floods in the river Bhima at Deorgaon, a tributary of the river Krishna, for the period 1951 to 1977 is given below. Verify whether the Gumbel extreme-value distribution fit the recorded values. Estimate the flood discharge with recurrence interval of (i) 100 years and (ii) 150 years by graphical extrapolation.*

Year	1951	1952	1953	1954	1955	1956	1957	1958	1959
Max. flood (m^3/s)	2947	3521	2399	4124	3496	2947	5060	4903	3757
Year	1960	1961	1962	1963	1964	1965	1966	1967	1968
Max. flood (m^3/s)	4798	4290	4652	5050	6900	4366	3380	7826	3320
Year	1969	1970	1971	1972	1973	1974	1975	1976	1977
Max. flood (m^3/s)	6599	3700	4175	2988	2709	3873	4593	6761	1971

SOLUTION: The flood discharge values are arranged in descending order and the plotting position recurrence interval T_p for each discharge is obtained as

$$T_p = \frac{N+1}{m} = \frac{28}{m}$$

where m = order number. The discharge magnitude Q are plotted against the corresponding T_p on a Gumbel extreme probability paper (Fig. 1).

The statistics $\bar{x}$ and σ_{n-1} for the series are next calculated and are shown in Table 3 Using these the discharge x_T for some chosen recurrence interval is calculated by using Gumbel's formulae [Eqs. (5), (8) and (7)].

TABLE 3 – CALCULATION OF T_p FOR OBSERVED DATA
EXAMPLE 1

Order number m	Flood discharge x (m^3/s)	T_p (years)	Order number m	Flood discharge x (m^3/s)	T_p (years)
1	7826	28.00	15	3873	1.87
2	6900	14.00	16	3757	1.75
3	6761	9.33	17	3700	1.65
4	6599	7.00	18	3521	1.56
5	5060	5.60	19	3496	1.47
6	5050	4.67	20	3380	1.40
7	4903	4.00	21	3320	1.33
8	4798	3.50	22	2988	1.27
9	4652	3.11	23	2947	—
10	4593	2.80	24	2947	1.17
11	4366	2.55	25	2709	1.12
12	4290	2.33	26	2399	1.08
13	4175	2.15	27	1971	1.04
14	4124	2.00			

N = 27 years, $\bar{x}$ = 4263 m^3/s, σ_{n-1} = 1432.6 m^3/s

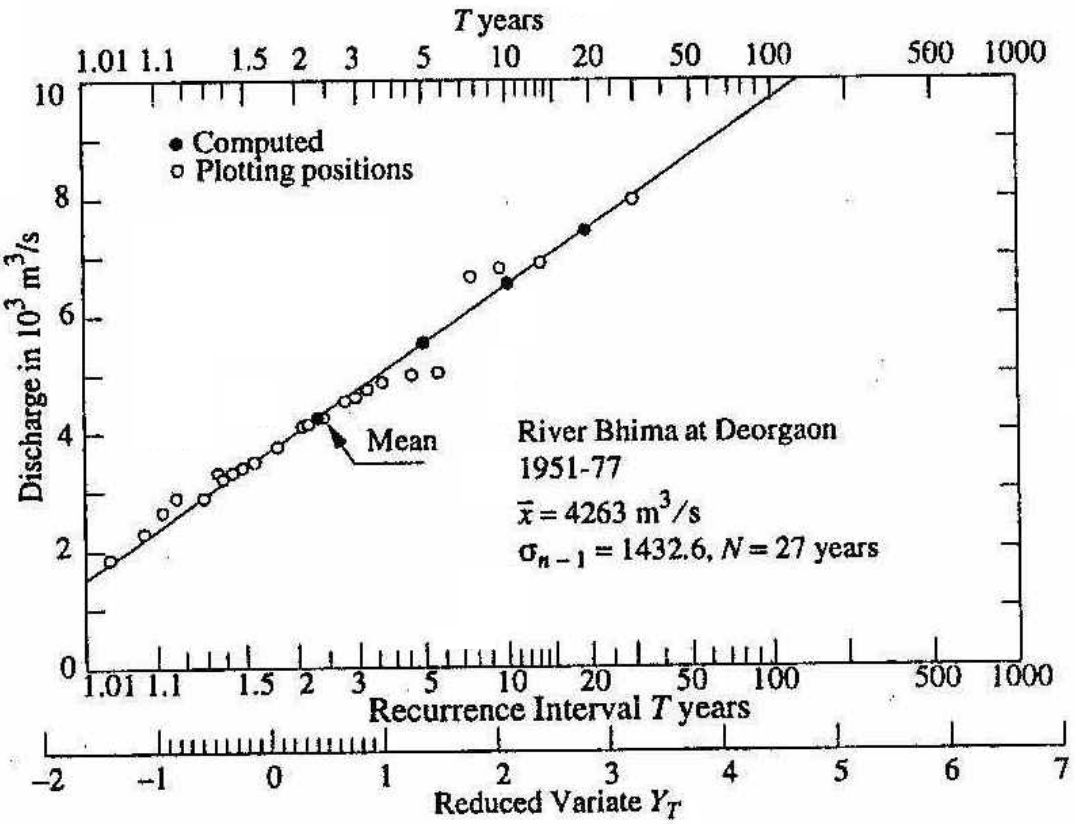

Fig. 1 – Flood probability analysis by Gumbel's distribution

From Tables-1 and-2, for $N = 27$, $y_n = 0.5332$ and $S_n = 1.1004$.
Choosing $T = 10$ years, by Eq. (5),

$$y_T = -[\ln . \ln (10/9)] = 2.25037$$

$$K = \frac{2.25307 - 0.5332}{1.1004} = 1.56$$

$$\bar{x}_T = 4263 + (1.56 \times 1432.6)$$

$$= 6499 \text{ m}^3/\text{s}$$

Similarly, values of x_T are calculated for two more T values as shown below.

T (years)	x_T [obtaained by Eq. (7.)] (m^3/s)
5.0	5522
10.0	6499
20.0	7436

These values are shown in Fig. 1 . It is seen that due to the property of the Gumbel's extreme probability paper these points lie on a straight line. A straight line is drawn through these points. It is seen that the observed data fit well with the theoretical Gumbel's extreme-value distribution.

[*Note*: In view of the linear relationship of the theoretical x_T and T on a Gumbel probability paper it is enough if only two values of T and the corresponding x_T are calculated. However, if Gumbel's probability p per is not available, a semi-log plot with log scale for T will have to be used and a large set of (x_T, T) values are needed to identify the theoretical curve.]

By extrapolation of the theoretical x_T vs T relationship, from Fig. 1 ,

At $T = 100$ years, $x_T = 9600$ m^3/s

At $T = 150$ years, $x_T = 10{,}700$ m^3/s

[By using Eq. (7), (8) and (5) , $x_{100} = 9558$ m^3/s and $x_{150} = 10088$ m^3/s.]

EXAMPLE-2 . *Flood-frequency computations for the river Chambal at Gandhisagar dam, by using Gumbel's method, yielded the following results:*

Return period T (years)	Peak flood (m^3/s)
50	40,809
100	46,300

Estimate the flood magnitude in this river with a return period of 500 years.

SOLUTION : By Eq. (7),

$$x_{100} = \bar{x} + K_{100}\,\sigma_{n-1}$$

$$x_{50} = \bar{x} + K_{50}\,\sigma_{n-1}$$

$$(K_{100} - K_{50})\,\sigma_{n-1} = x_{100} - x_{50} = 46300 - 40809 = 5491$$

But
$$K_T = \frac{y_T}{S_n} - \frac{\bar{y}_n}{S_n}$$

where S_n and $\bar{y}_n$ are constants for the given data series.

$$\therefore \qquad (y_{100} - y_{50})\frac{\sigma_{n-1}}{S_n} = 5491$$

By Eq. (5),

$$y_{100} = -[\ln . \ln (100/99)] = 4.60015$$

$$y_{50} = -[\ln . \ln (50/49)] = 3.90194$$

$$\frac{\sigma_{n-1}}{S_n} = \frac{5491}{(4.60015 - 3.90194)} = 7864$$

For $T = 500$ years, by Eq. (5),

$$y_{500} = -[\ln . \ln (500/499)] = 6.21361$$

$$(y_{500} - y_{100})\frac{\sigma_{n-1}}{S_n} = x_{500} - x_{100}$$

$$(6.21361 - 4.60015) \times 7864 = x_{500} - 46300$$

$$x_{500} = 58988, \quad \text{say } 59{,}000 \text{ m}^3/\text{s}$$

Confidence Limits

Since the value of the variate for a given return period, x_T determined by Gumbel's method can have errors due to the limited sample data used, an estimate of the confidence limits of the estimate is desirable. The confidence interval indicates the limits about the calculated value between which the true value can be said to lie, with a specific probability based on sampling errors only.

For a confidence probability c, the confidence interval of the variate x_T is bounded by values x_1 and x_2 given by

$$x_{1/2} = x_T \pm f(c)\, S_e \qquad \text{.................................(9)}$$

where $f(c)$ = function of the confidence probability c determined by using the table of normal variates as

c in per cent	50	68	80	0	95	99
$f(c)$	0.674	1.00	1.282	1.645	1.96	2.58

$$S_e = \text{probable error} = b\,\frac{\sigma_{n-1}}{\sqrt{N}} \qquad \text{...............................(9-a)}$$

$$b = \sqrt{1 + 1.3\,K + 1.1\,K^2}$$

K = frequency factor give by Eq. (8)

σ_{n-1} = standard deviation of the sample

N = sample size.

It is seen that for a given sample and T, 80% confidence limits are twice as large as the 50% limits and 95% limits are thrice as large as 50% limits.

EXAMPLE –3 *Data covering a period of 92 years for the river Ganga at Raiwala yielded the mean and standard derivation of the annual flood series as 6437 and 2951* m^3/s *respectively. Using Gumbel's method estimate the flood discharge with a return period of 500 years. What are the (a) 95% and (b) 80% confidence limits for this estimate.*

SOLUTION: From Table – 1 for N = 92 years, $\bar{y}_n = 0.5589$ and $S_n = 1.2020$ from Table – 2

$$y_{500} = -[\ln . \ln (500/499)]$$

$$= 6.21361$$

$$K_{500} = \frac{6.21361 - 0.5589}{1.2020} = 4.7044$$

$$x_{500} = 6437 + 4.7044 \times 2951 = 20320 \text{ m}^3/\text{s}$$

From Eq. (9 a)

$$b = \sqrt{1 + 1.3\,(4.7044) + 1.1\,(4.7044)^2}$$

$$= 5.61$$

$$S_e = \text{probable error} = 5.61 \times \frac{2951}{\sqrt{92}} = 1726$$

(a) For 95% confidence probability $f(c) = 1.96$ and by Eq. (9)

$$x_{1/2} = 20320 \pm (1.96 \times 1726)$$

$$x_1 = 23703 \text{ m}^3/\text{s} \quad \text{and} \quad x_2 = 16937 \text{ m}^3/\text{s}$$

Thus estimated discharge of 20320 m^3/s has a 95% probability of lying between 23700 and 16940 m^3/s

(b) For 80% confidence probability, $f(c) = 1.282$ and by Eq. (9)

$$x_{1/2} = 20320 \pm (1.282 \times 1726)$$

$$x_1 = 22533 \text{ m}^3/\text{s} \quad \text{and} \quad x_2 = 18107 \text{ m}^3/\text{s}$$

The estimated discharge of 20320 m^3/s has a 80% probability of lying between 22530 and 18110 m^3/s.