Training Course

On

Predictions in Ungauged Basins

[July 26 – 30, 2010]

CHAPTER-4

FLOOD FREQUENCY ANALYSIS TECHNIQUES

By

Rakesh Kumar,
Scientist-F

Organized under Hydrology Project - II

National Institute of Hydrology
Roorkee – 247667 (Uttrakhand)

CHAPTER 4

FLOOD FREQUENCY ANALYSIS TECHNIQUES

Rakesh Kumar
Scientist-F, Head (Surface Water Div.)
National Institute of Hydrology, Roorkee–247 667

4.0 INTRODUCTION

In many problems in hydrology, the data consists of measurements on a single random variable; hence we must deal with univariate analysis and estimation. The objective of univariate analysis is to analyze measurements on the random variable, which is called sample information, and identify the statistical population from which we can reasonably expect the sample measurements to have come. After the underlying population has been identified, one can make probabilistic statements about the future occurrences of the random variable, this represents univariate estimation. It is important to remember that univariate estimation is based on the assumed population and not the sample, the sample is used only to identify the population.

Hydrologic processes such as rainfall, snowfall, floods, droughts etc. are usually investigated by analysing their records of observations. Many characteristics of these processes may not represent definite relationship. For example, if you plot instantaneous peak discharges from each year for a river, a rather erratic graph is obtained. The variation of peak discharge from one year to another can not be explained by fitting a definite relationship, which we call as deterministic relationship. For the purpose of hydrologic analysis, the annual peak discharge is then considered to be a random variable. Methods of probability and statistics are employed for analysis of random variables. In this lecture, some elementary concepts of probability and statistics are presented, which are used for frequency analysis in hydrology.

The following assumptions are implicit in frequency analysis in order to have meaningful estimates from flood frequency analysis:

- The data to be analysed describe random events.
- It is homogeneous.
- The population parameters can be estimated from the sample data.
- It is of good quality.

If the data available for analysis do not satisfy any of the above listed assumptions then, much reliability can not be attached to the estimates. For flood frequency analysis, either annual flood series or partial duration series may be used. The requirements with regard to data are that,

- it should be relevant,
- it should be adequate and,
- it should be accurate.

In general, an array of annual peak flood series may be considered as a sample of random and

independent events. The non-randomness of the peak series will, however, increases the degree of uncertainty in the derived frequency relationship. Various tests are available to check the randomness of the peak flow data. The annual maximum flood series can generally be regarded as consisting of random events as the mean interval of each observed flood peak is 1 year. However, in the case of data used for partial duration series analysis, the independence among the data is doubtful. The peaks are selected in such a way that they constitute a random sample.

The term relevant means that data must deal with problem. For example, if the problem is of duration of flooding then data series should represent the duration of flows in excess of some critical value. If the problem is of interior drainage of an area then data series must consist of the volume of water above a particular threshold. The term adequate primarily refers to length of data. The length of data primarily depends upon variability of data and hence there is no guideline for the length of data to be used for frequency analysis.

The term accurate refers primarily to the homogeneity of data and accuracy of the discharge figures. The data used for analysis should not have any effect of man made changes. Changes in the stage discharge relationship may render stage records non-homogeneous and unsuitable for frequency analysis. It is therefore preferable to work with discharges and if stage frequencies are required then most recent rating curve is used. Watershed history and flood records should be carefully examined to ensure that no major watershed changes have occurred during the period of record. Only records, which represent relatively constant watershed conditions, should be used for frequency analysis. The fundamental terms related basic statistics are explained in the earlier lecture note. In the following section some of the popular methods used in flood frequency analysis are explained with suitable examples.

4.1 PROBABILITY DISTRIBUTION

A distribution is an attribute of a statistical population. If each element of a population has a value of X then the distribution describes the constitution of the population as seen through its X values. It tells whether they are in general very large or very small, that is their location on the axis. It tells whether they are bunched together or spread out and whether they are symmetrically disposed on the X axis or not. These three things are described by the mean, standard deviation and skewness.

Distribution also tells the relative frequency or proportion of various X values in the population in the same way that a histogram gives that information about a sample. These relative frequencies are also probabilities and hence the distribution tells us the probability, $Pr(X<x)$, that the X value on an element drawn randomly from the population would be less than a particular value x. Knowing $Pr(X<x)$ for all x values, the laws of probability may then be used to deduce the probability of any proposition about the behavior of a random sample of X values drawn from the population.

When the population is sufficiently large the histogram of its X values can be made with very small class intervals and the histogram can be replaced by a smooth curve, the area enclosed by any two vertical ordinates being the relative frequency or probability of x values between those ordinates. Because of this probability interpretation, a relative frequency distribution is also called a probability distribution and the curve describing it is called a probability density function (p.d.f) whose cumulative function is called the distribution function (d.f.). In flood frequency analysis, the sample data is used to fit probability distribution which in turn is used to extrapolate from recorded events to design events either graphically or analytically by estimating the parameters of the distribution. A large number of peak flow distributions are available in literature.

4.1.1 Normal Distribution

The normal distribution is one of the most important distribution in statistical hydrology. This is a bell shaped symmetrical distribution having coefficient of skewness equal to zero. The normal distribution enjoys unoque position in the field of statistics due to central limit theorem. This theorem states that under certain very-broad conditions, the distribution of sum of random variables tends to a normal distribution irrespective of the distribution of random variables, as the number of terms in the sum increases. The Probability Density Function (PDF) and Cumulative Density Function (CDF) of the distribution are given in Appendix-I.

4.1.2 Log Normal Distribution

The causative factors for many hydrologic variables act multiplicatively rather than additively and so the logarithms of these variables which are the product of these causative factors follow the normal distribution.

If $Y = \log_e(X)$ follows normal distribution, the X is said to follow log normal distribution. If the variable X has a lower bound X_0, different from zero and the variable $Y = \log_e(X-X_0)$ follows normal distribution then X is log normally distributed with three parameters. The PDF and CDF of the distribution are given in Appendix.

4.1.3 Pearson type-III Distribution (PT3)

Pearson type III distribution is a three parameter distribution. This is also known as Gamma distribution with three parameters. The PDF and CDF of the diatribution are given in Appendix-I.

4.1.4 Exponential Distribution

Exponential distribution is a special case of Pearson type III distribution when shape parameter $\gamma = 1$. The PDF and CDF are given in appendix-I.

4.1.5 Gumbel Extreme Value (Type-I) Distribution (EV1)

One of the most commonly used distributions in flood frequency analysis is the double exponential distribution (known as Gumbel distribution or extreme value type I or Gumbel EV1 distribution). The PDF and CDF of the distribution are given in Appendix-I.

4.1.6 Log Pearson Type-III Distribution (LP3)

If $Y = \log_e(X)$ follows Pearson type III distribution then X is said to follow log Pearson type III distribution. In 1967, the U.S. water Resources Council recommended that the log Pearson type III distribution should be adopted as the standard flood frequency distribution by all U.S. federal government agencies. The PDF and CDF of the distribution are given in Appendix-I.

4.1.7 Extreme Value Distribution

Just as there is a family of Pearson type III distributions, each member being characterized by a value of γ, there is also a family of EV distributions, each member of which is characterized by the value of a parameter denoted by k. The family can be divided into three classes, corresponding to different ranges of k values. The three classes are referred as Fisher Tippett type 1, type2 and type 3. They are also known as EV-1, EV-2 and EV-3 distributions. In practice, k value lie in the range -0.6 to + 0.6. For EV-1 distribution k is zero and coefficient of skewness is equal to 1.139. For EV-2 distribution, the values of k is -ve and skewness is greater than 1.139. Foe EV3 distribution the value of k is +ve and coefficient of skewness is less than 1.139. EV-1 and EV-2 distributions are known as Gumbel and Frechet distributions respectively.

4.2 METHODS OF PARAMETER ESTIMATION

There are four well known parameter estimation techniques viz:

1. Graphical
2. Least squares
3. Method of moments , and
4. Method of maximum likelihood.
5. Method of Probability Weighted Moment .

4.2.1 Graphical Method

In graphical method of parameter estimation, the variate under consideration is regarded as a function of a reduced variate of a known distribution. The steps involved in the graphical method are as follows:

1. Arrange the variates of annual maximum flood series in ascending order and assign different ranks to the individual variates.
2. Assign the plotting positions to each of the variates. The plotting position formula may be used depending upon the type of distribution being fitted. The recommended plotting position formulae for normal, log normal, Gumbel EV-I, Pearson type-III and Log Pearson type-III distributions are given in Table. 4.1. Where 'm' indicates the rank number, m=1 for the smallest observation and m= N for largest observation when flood series is arranged in ascending order.
3. Estimate the reduced variates for the selected Distribution corresponding to different plotting positions, which represent the probability of non-exceedence.

The reduced variates for normal and log normal distributions are computed with the help of table 2. For Gumbel (EV1) Distribution, the reduced variates are computed using eq. (15) of Appendix-I, where F(z) represents probability of non-exceedence which have been computed using the suitable plotting position formula, and z is the corresponding reduced variates. In case of Pearson type-III and Log Pearson type-III which are three parameter distributions different sets of reduced variates are obtained for different coefficients of skewness.

At this stage it is essential to introduce the concept of frequency factor as it has been found easier to develop probability paper based on the frequency factor, K for Pearson type-III and Log Pearson type-III distributions on comparison to that developed based on the reduced variate

Table 4.1 Unbiased Plotting Position Formulae

Distributions(s)	**Recommended plotting position formula**	**Form of the plotting position formulae**
Normal and Log Normal	Blom	$F(X\#x) = \frac{(m-3/8)}{(N+1/4)}$
Gumbel EV-I	Gringorton	$F(X\#x) = \frac{(m-0.44)}{(N+0.12)}$
Pearson type III and Log Pearson type III	Cunnane	$F(X\#x) = \frac{(m-0.40)}{(N+0.20)}$

Frequency Factor

A general frequency equation proposed by Chow, applicable for different distributions is in the following form:

$X = m + Ks$ (1)

where

X = Magnitude of the different variates of the peak flood series, or the magnitude of the flood at required return period T.

K = Frequency factor corresponding to X. For Pearson Type-III and Log Pearson Type-III distribution, the values of frequency factors are given in appendix-II at different probability of exceedence levels corresponding to the various coefficients of skewness values. Where, m and s = Mean and standard deviation of the population which would be replaced by the sample statistics.

The step by step procedure is as follows :

1. Plot the sample data as a series of discrete points on an ordinary graph paper with the ordinate being the variate and the abscissa being the reduced variate or frequency factor. Such plot can be prepared for different distribution.
2. Draw a best-fit line through the plotted points. The slope and intercept of the line provide the

estimates for the parameters of the respective distributions. This straight line can be projected to arrive at the flood magnitudes of desired return periods.

In graphical estimation procedure the line is subjectively placed and could vary with analyst. This subjectivity is regarded as a major drawback by hydrologists. The following example provides the procedure for estimating the reduced variate corresponding to a given probability level for Normal, EV1 and Pearson Type- III distributions:

Example 4.1:
(i) Find out the value of the reduced variates for Normal and EV1 distributions corresponding to probability of non-exceedence of 0.30 from Appendix-II.

(ii) Find out the frequency factor K for Pearson Type-III distribution corresponding to non-exceedence probability of 0.30 and coefficient of skewness of 1.3 from Appendix-III.

Solution:
(i) For Normal distribution use table given as appendix-II to estimate the reduced variate corresponding to probability of non-exceedence equal to 0.30. In this table, areas under the normal curve are given. Since the probability of non-exceedence, F(z) is 0.3 which is less than 0.5, the value of Z will be a negative quantity.

Find out a value of Z' which has non-exceedance probability=0.30.
Z' = -0.5244.

The reduced variate corresponding to F(Z)=0.3 can be obtained for EVI distribution using equation given in Appendix-I.

$$0.30 = e^{-e^{-Z}}$$

$$Z = -\ln(-\ln(0.30)) = -0.186$$

(ii) Table given as Appendix-III may be used to estimate the frequency Factor, K for Pearson Type III distribution.

Probability of exceedence = 1 - 0.30 = 0.70
and coefficient of skewness=1.3
Frequency Factor, K = -0.634. (From the table of appendix-III)

4.2.2 Least Squares Method

In the method of least squares for the parameter estimation the steps from (i) to (ii) of the Graphical method may be repeated. In this technique a simple linear regression equation is fitted between the variate under consideration and the corresponding reduced variate or frequency factor, rather than drawing a best fit line subjectively on a simple graph paper. This method has not been accepted as a standard method in practice as it involves the use of plotting position formula to determine the reduced variate or the frequency factor and due to the assumption that the error variance remains same for all observations. The defect due to the former assumption could be eliminated by using the appropriate plotting positions formula given in table 1. However, the later assumption makes the method more defective as the higher events recorded have more error variance than the recorded lower events. All these assumptions affect the correct estimation of the slope and intercept of the line, which represent the parameters of the distribution.

4.2.3 Method of Moments

The method of moments makes use of the fact that if all the moments of a distribution are known then everything about the distribution is known. For all the distributions in common usage, four moments or fewer are sufficient to specify all the moments. For instance, two moments, the first together with any moment of even order are sufficient to specify all the moments of the Normal Distribution and therefore the entire distribution. Similarly, in the Gumbel EV Type-I Distribution, the first two moments are sufficient to specify all the moments and hence the distribution. For Pearson Type III Distribution three moments, always taken as the first three required to specify all the moments. In these cases the number of moments needed to specify all the moments and hence the distribution equals the number of parameters.

The method of moment's estimation is dependent on the assumption that the distribution of variate values in the sample is representative of the population distribution. Therefore, a representation of the former provides an estimate of the later. Given that the form of the distribution is known or assumed, the distribution which the sample follows is specified by its first two or three moments calculated from the data.

4.2.4 PROBABILITY WEIGHTED MOMENT (PWM) AND LINEAR (L) MOMENTS

Greenwood et al. (1979) introduced the method of probability weighted moment (PWM) and showed its usefulness in deriving explicit expressions for the parameters of the distributions whose inverse forms x = x (F) can be explicitly defined. If probability distribution function be denoted by F = F(x) = P[X ≤ x], the PWM's are the moment of the function x (F) and is expressed as,

$$M_{i,j,k} = E[x^i F^j (1-F)^k = \int_0^1 [x(F)]^i F^j (1-F)^k \, dF \qquad (2)$$

It can be shown that,

$$M_{i,0,k} = \sum_{j=0}^{k} (c_j^k)(-1)^j M_{i,j,0} \qquad (3)$$

Where $M_{I,j,k}$ is the PWM of order of i, j and k and E is the expectation operator, then $M_{I,0,0}$ represents the conventional moment about the origin of order 'i'. These can than be related to the distribution parameters and the resultant relation may be a simpler structure than conventional moments and the parameters. For example the parameters of Gumbel (EV1) distribution using PWM's is as follows;

$$x(F) = u - \sigma \, Ln[-Ln(F)] \quad (4)$$

$$M_{i,j,0} = \int_0^1 [u - \sigma \ln(-\ln F)]^i F^j (1-F)^0 dF \quad (5)$$

$$= \int_0^1 [uF^j] dF - \int_0^1 \sigma \ln(-\ln F)]^i F^j dF \quad (6)$$

$$= \frac{u}{1+j} + \sigma \frac{\ln(1+j) + \varepsilon}{1+j} \quad (7)$$

$$M_{1,0,0} = u + \sigma\varepsilon \quad (8)$$

$$M_{1,1,0} = \frac{u}{2} + \sigma\left[\frac{Ln(2) + \varepsilon}{2}\right] \quad (9)$$

Solving (7) and (8),, $\sigma = \dfrac{M_{1,0,0} - 2M_{1,0,1}}{Ln(2)}$ (10)

And $u = M_{1,0,0} - 0.5772\sigma$, $\varepsilon = 0.5772$ (11)

L-moments are linear combination of order statistics which are robust to outliers and are unbiased for small samples, making them suitable for flood frequency analysis, including identification of distribution and parameter estimation (Hosking,1990, Hosking and Wallis, 1993). L-moments are identified as linear combination of probability weighted moment (PWM)

$$M_{1,r,0} = \beta_r = E\{X[F(x)]^r\} \quad (12)$$

where F(x) is the cumulative distribution function of x. When r=0, β_0 is the mean. The first four L-moments expressed as linear combination of PWM's are :

$$\lambda_1 = \beta_0$$
$$\lambda_2 = 2\beta_1 - \beta_0$$
$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0$$
$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \quad (13)$$

Landwehr et al (1979) recommends the use of biased estimates of PWM's and L-moments, since such estimates often produces quintile estimates with lower root mean square error than unbiased alternatives. Nevertheless, unbiased estimators are preferred in moment diagrams for goodness of fit for less bias and are invariant when the data is multiplied by a constant. Unbiased estimates of PWM's for any distribution can be computed as follows,

$$b_0 = \frac{1}{n}\sum_{j=1}^{n} x_j\ ; b_1 = \sum_{j=1}^{n-1}\left[\frac{(n-j)}{(n)(n-1)}\right]x_j ; b_2 = \sum_{j=1}^{n-2}\left[\frac{(n-j)(n-j-1)}{n(n-1)(n-2)}\right]x_j ;$$

$$b_3 = \sum_{j=1}^{n-3}\left[\frac{(n-j)(n-j-1)(n-j-2)}{n(n-1)(n-2)(n-3)}\right]x_j \qquad (14)$$

where x_j represents the ordered streamflow with x_1 the largest streamflow and x_n the smallest. The sample estimates of L-moments are calculated by replacing b_0, b_1, b_2 and b_3 for $\beta_0, \beta_1, \beta_2$ and β_3 in equation (12). So the parameters of Gumbel (EV1) in the above example can be written in terms of L-moment as follows,

$$\sigma = \frac{M_{1,0,0} - 2M_{1,0,1}}{Ln(2)} = -2(\lambda_2)/\ln(2) \qquad (15)$$

$$u = \lambda_0 - 0.5772\sigma$$

The relationship equations between sample moment parameters and population parameter estimation for different distributions using L-moment are given in Annexure-I for ready reference. Although the theory and application of L-moment parallels those of conventional moment , there are certain advantages of L-moment.Since sample estimators of L-moments are always combination of ranked observations, they are subjected to less bias than ordinary product moments. This is because ordinary product moment require squaring and cubing giving greater weightage to observations far from the mean resulting in bias.

4.3 RETURN PERIOD FLOOD ESTIMATION

4.3.1 Normal Distribution

The parameters of the normal distribution, μ and σ, which describe the characteristics of the given set, are computed as:

$$\mu \equiv \overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \qquad (16)$$

$$\sigma \equiv s = \frac{\sqrt{\Sigma(X_i - \overline{X})^2}}{(N-1)} \qquad (17)$$

After computing the parameters, μ and σ, the T-year flood estimates can be obtained using the Normal Distribution. The steps are:

1. Compute the $\overline{X}$ and s from the sample data. $\overline{X}$, and s provide estimates for the parameter, μ and σ.

2. Compute the probability of non-exceedence using the relation:

$$F(X < x) = 1 - 1/T \qquad (18)$$

3. Compute the Normal Reduced Variate Z_T from the statistical table , corresponding to the probability of non-exceedence computed at step (2).

4. Estimate the flood for T-year recurrence interval using the following equation:

$$X_T = \mu + \sigma Z_T \tag{19}$$

4.3.2 Log Normal Distribution (two parameters)

In log normal two parameter distribution the variates are transformed to the log domain by taking log of each variate, and then the mean and standard deviation of the transformed series are computed as given by eq. (1) and (2) of lecture note on basic statistics. The flood at required recurrence interval can be computed in the following steps after fitting the log normal distribution with the sample data.

1. Transform the original series of peak flood data into log domain by taking log of each variate to the base e.
2. Compute the mean, $\overline{Y}$ and standard deviation, S_Y, from the log transformed series. $\overline{Y}$ and S_Y provide estimates for the parameters, μ_Y and σ_Y , respectively.
3. Compute the probability of non-exceedance for the given recurrence interval (T).
4. Compute the normal reduced variate, Z_T corresponds to the probability of non-exceedance computed at step (3).
5. Estimate the flood for T-year recurrence interval in log domain using the following equation:
$$Y_T = \mu_Y + \sigma_Y Z_T \tag{20}$$
6. Transform the estimated T-year flood in original domain by computing its exponent i.e.
$$X_T = e^{Y_T} \tag{21}$$

4.3.3 Gumbel's Extreme value type-I Distribution

The relationship between the parameters and the statistical moments of the data are given by

$$\mu = u + 0.5772\ \alpha \tag{22}$$
$$\alpha = \pi / \alpha^2\ (6)^{1/2} = 1.645 / \alpha^2 \tag{23}$$

Solving equations (28) and (29) for u and α, we get

$$\alpha = 0.7797\ \sigma \tag{24}$$
$$u = \mu - 0.45\ \sigma \tag{25}$$

The population statistics, μ and σ would be replaced by the sample statistics, $\overline{X}$ and respectively. The step by step procedure for computing the T-year flood using EV1 distribution is given below:

1. Compute mean, $\overline{X}$ and standard deviation, S, from the sample. $\overline{X}$, and S provide estimates for μ and σ respectively.

2. Compute the parameter u and α using eq. (30) and (31).
3. Compute the probability of non-exceedence F(Z) corresponding to T-year recurrence interval
4. Compute the EV1-reduced variate, Z_T, using the relationship;

$$Z_T = - \ln(-\ln F(Z)) \qquad (26)$$

5. Estimate T-year recurrence interval flood using the EV1 distribution as follows:

$$X_T = u + \alpha Z_T \qquad (27)$$

4.3.4 Pearson Type -III Disribution (PT-III)

PT-III distribution is a three parameter distribution. Three moments are, therefore, needed for computing the parameters. Mean, standard deviation and skewness computed from the sample data describe the measures for first three moments of the sample data. The following steps are usually involved in computing the T-year flood using the Pearson Type-III Distribution.

1. Compute the mean, $\overline{X}$ and standard deviation S, using eq (1) and (2) respectively. Compute the coefficient of skewness g from the sample using the following equation:

$$g = \frac{N}{(N-1)(N-2)} \sum_{i=1}^{N} \frac{(X_i - \overline{X})^3}{S} \qquad (28)$$

2. Compute the probability of exceedence for the given recurrence interval, T, which equals to 1/T.
3. Estimate the frequency factor, K_T from the table corresponding to the computed coefficient of skewness g, and the probability of exceedence.
4. Estimate T-year flood using the equation;

$$X_T = \overline{X} + S\,K_T \qquad (29)$$

Example 4.2:

The mean, standard deviation and coefficient of skewness of original and log transformed annual maximum peak flood series of a typical gauging site are given below:

	Original Series	Log transformed series
Mean(m^3/s)	506.843	6.157
Standard dev.(m^3/s)	211.087	0.372
Coefficient of skewness	1.564	0.556

Estimate 1000-years floods assuming that the peak discharge data follow:(i) Log Normal Distribution, (ii) Log Pearson Type-III, (iii) Pearson Type-III and (iv) Gumbel EV1 Distribution.

Solution:

(i) Flood estimate Assuming Log Normal Distribution

Since μ_Y / $\overline{Y}$ = 6.157
σ_Y / S_Y = 0.372 and,
Probability of non-exceedence, F (Z) = 1 - 1/1000 = 0.999

^ Z_T = 3.10 for T = 1000 years (using Statistical Table)
^ $Y_T = \mu_Y + \sigma_Y Z_T$
Y_{1000} = 6.157 + 0.372 x 3.10 = 7.3102
^ $X_{1000} = e^{Y}1000 = e^{7.3102}$ = 1495 m^3/s

(ii) Flood Estimate Assuming Log Pearson Type -III Distribution
Since μ_Y / $\overline{Y}$ = 6.157
σ_Y / S_Y = 0.372 and,
γ_Y / g_Y = 1.564

Probability of exceedance 1-F(Z) = 1/1000 =0.001
Frequency Factor, K_T = 3.8919712 (using table)

^ $Y_T = \mu_Y + \sigma_Y K_T$

Y_{1000} = 6.157 + 0.372 x 3.8919712 = 7.6048
$X_{1000} = e^{7.6048}$ = 2007.8 m^3 /s

(iii) Flood estimate assuming Pearson type-III Distribution
Since μ / $\overline{X}$ = 506.843
σ / S = 211.087 and,
g / g_X = 1.564

Probability of exceedence, 1-F(Z) = 1/1000 = 0.001
Frequency factor K_T = 5.3214276 (using table given in Appendix-III)

^ $X_T = \mu + \sigma K_T$
= 506.843+211.08 x 5.3214276
= 1630 m^3/s

(iv) Flood Estimate Assuming Gumbel EV1 Distribution
Since μ / $\overline{X}$ = 506.843, σ / S = 211.087 and,

α = 0.7797 x σ = 0.7797 x 211.087 = 164.585
u = μ - 0.45 σ
u = 506.843 - 0.45 x 211.087
= 411.854
Z_T = - ln (- ln (F (Z))) ; F(Z) =1 - 1/1000 ; Z_{1000} = - ln (-ln (0.999)) = 6.907255

$X_{1000} = u + \alpha Z_{1000}$; ^ $X_{1000} = 411.854 + 164.585 \times 6.907255 = 1548.68\ m^3/s$

4.4. TEST GOODNESS OF FIT

The validity of a probability distribution function proposed to fit the frequency distribution of a given sample may be tested graphically or by analytical methods. Graphical approaches are usually based on comparing visually the probability density function with the corresponding empirical density function of the sample under consideration. In other words model CDF is compared with empirical CDF. Often these CDF graphs are made on specifically designed paper such that the model CDF plots as a straight line. An example of this is the Gumbel paper. If empirical CDF plots as a straight line on the Gumbel paper it is an indication that the Gumbel distribution may be valid model for the data at hand. Often graphical approaches for judging how good a model is, are quite subjective.

A number of analytical tests have been proposed for testing the goodness of fit of proposed models. Some of these tests are presented here under.

(a) Coefficient of determination (r^2): It describes the extent of best-fit and is expressed as:

$$r^2 = \left[\Sigma(x_i - \bar{x})\ (y_i - \bar{y})\right]^2 / [\Sigma(x_i - \bar{x})^2 \Sigma(y_i - \bar{y})^2] \qquad (30)$$

(b) Coefficient of correlation (r): It is the square root of the coefficient of determination (r^2):

$$r = \sqrt{r^2} = \Sigma(x_i - \bar{x})(y_i - \bar{y}) / [\Sigma(x_i - \bar{x})^2 \Sigma(x_i - \bar{y})^2]^{1/2} \qquad (31)$$

where 'x' is the observed and 'y' is the computed value.

(c) Efficiency (EFF): The efficiency of the best-fit is given as (Ref.*****)

$$EFF = 1 - \frac{S}{S_y} \qquad (32)$$

where,

$$S^2 = \Sigma(y_i - \hat{y}_i)^2/(n-2) \qquad (33)$$

$$S_y^2 = \Sigma(y_i - \bar{y})^2/(n-1) \qquad (34)$$

where 'n' is the total number of observations.

(d) Standard error (SE): Standard errors of the estimated regression coefficients a and b are computed, respectively as:

$$S_a = S\left(\frac{1}{n} + \frac{\bar{x}^2}{\Sigma(x_i - \bar{x})^2}\right)^{1/2} \qquad (35)$$

$$S_b = S/\ \Sigma(x_i - \bar{x}^2)^{1/2} \qquad (36)$$

where S_a and S_b are standard errors of coefficients a and b, respectively.

(e) Confidence interval: The confidence interval for a is given as:

$$l_a = a - t_{(1-\alpha/2),\ (n-2)}\ S_a \qquad (37)$$

$u_a = a + t_{(1-\alpha/2),\,(n-2)}\, S_a$ (38)

and for b, it is given as:

$l_b = b - t_{(1-\alpha/2),\,(n-2)}\, S_b$ (39)

$u_b = b + t_{(1-\alpha/2),\,(n-2)}\, S_b$ (40)

where l_a, u_a and l_b, u_b denote lower and upper confidence limits of a and b, respectively; α is the confidence level; and $t_{(1-\alpha/2),\,(n-2)}$ represent t-values corresponding to (1- α/2) confidence limits and (n-2) degrees of freedom. These t-values can be obtained from the table given in Appendix III.

4.5 REGIONAL FLOOD FREQUENCY ANALYSIS

4.5.1 Index Flood Method :

Basically, the Index-Flood method (Dalrymple, 1960) extrapolates statistical information of runoff events for flood frequency analysis from gauged catchments to ungauged catchments in the vicinity having similar catchment & hydrologic characteristic. The follwing steps are given in sequential steps for estimating return period flood using the above method,

1. Select the gauged catchments within the region having similar characteristic to the ungauged catchments.
2. Determine the base period to be used for the study.
3. Establish flood frequency curve is prepared with the ranked annual series and the corresponding plotting position is estimated using the Gringortan plotting formula.
4. The $Q_{2.33}$ (annual mean flood) is estimated using EV1 parameters.
5. Establish the relationship of mean annual flood and basin area.

6. Rank ratios of selected return periods to the mean annual flood at each site.
7. Compute median flood ratio for each of the selected return period of step (7), multiply by the estimated mean annual flood of the ungauged catchments and plot versus recurrence interval on Gumbel probability paper.

The last step is the flood frequency curves for ungauged catchments.

4.5.2 Method of L-Moments

Following sequential steps are followed,

1. Tests for regional homogeneity for the selected gauged catchments using the procedure described by Dalrymple (1960) & discard the catchments, which are not homogeneous.
2. Arrange the flood series and compute M_{100} , M_{101} and M_{102} using equations (19) and (20)
3. Standardize the computed values of M_{100}, M_{101}, and M_{102} obtained from step (ii) dividing them by the at site mean (same as M_{100}) . Hence:

$m_0 = M_{100} / M_{100} = 1$, $m_{i,j} = M_{101,j} / M_{100,j}$

4. Compute the regional value of the standardized PWM's averaged across the N's sites in the region in the ratio of the record lengths.

$m_i^{1} = \Sigma m_I\, l / (N)$

where m_I is the L-moment of the i th site and l is its record length. N is the sum total of all 'l'

5. Estimate the T- year recurrence interval flood using the relation (the pdf is optional):
   $X = u - [\ln[-\ln(1-1/T)]]$
6. Scale the quantiles X_T by at site mean in order to give an estimate for the site, Q
   $Q_T = M_{100} X_T$

For analysis using regional data the following steps are followed ,

1. Estimate the mean annual peak flood (Q_j) for each gauging site using the relationship between the mean annual peak floods and catchment area developed for the region.
2. Scale the quantile X_T by the mean obtained from the previous steps to estimate quantile

$Q_{T,j} = Q_j X_T$

4.6 REMARKS

The purpose of the frequency analysis is to estimate the design flood for desired recurrence interval assuming the sample data follow a theoretical frequency distribution. It is assumed that the sample data is a true representative of the population. It is generally seen that minimum 30 to 40 years of records are needed in order to carry out flood frequency analysis to the at site data for estimating the floods in extrapolation range, somewhat, within the desired accuracy. In case the length of records are too short, it represents inadequate data situation and at site flood frequency analysis fails to provide the reliable and consistent flood estimates. The regional flood frequency curves together with at site mean is generally able to provide more reliable and consistent estimates of floods under the inadequate data situation. For ungauged catchment, the regional flood frequency analysis approach is the only way to estimate the flood, for desired recurrence interval for which a regional relationship between mean annual peak flood and catchment characteristics is developed along with the regional frequency curves.

Most of the goodness of fit tests used for judging the best fit distribution represent the descriptive ability of the theoretical distributions. Generally the frequency distributions show larger deviations in extrapolation range. Thus a distribution which may pass the goodness of fit criteria not necessarily be able to estimate the floods for higher recurrence intervals to the desired accuracy. In order to judge the performance of the theoretical distributions in predicting the higher recurrence interval floods the predictive ability tests must be taken up using the Monte Carlo experiments with the data, generated by the selected distributions based on the descriptive ability criteria. Such generation studies provide a better opportunity for understanding the characteristics of the theoretical distributions. Those who are interested further in the area of flood frequency analysis may refer the relevant literatures appeared in national as well as international journals dealing with hydrology and water resources.

REFERENCES

1. Chow, V.T., 1964. *Hand book of Applied Hydrology.* Mc-Graw Hill New York.
2. Dalrymple, T (1960) 'Flood Frequency Methods', U. Survey Water Supply Paper 1543-A pp. 11~51, Washington.
2. Haan, C.T., 1977. *Statistical Methods in Hydrology.* The Iowa State University Press, U.S.A.
3. Kite, G.W., 1977. *Frequency and Risk Analysis in Hydrology.* Water Resources Publications,

Colorado.

4. Linseley, R.K., Kohler, H.A. and Paulhus, J.L.H., 1975. *Hydrology for Engineers.* Mc-Graw Hill, International Book Company.

5. McGuess, R.H. and Snyder, W.M., 1985. *Hydrologic Modelling, Statistical Methods and Applications.* Prentice-Hall, Englewood Cliffs, New Jersey.

6. Yevjevich,V., 1972. *Probability and Statistics Hydrology.* Water Resources Publications, Fort Collins, Colorado.

APPENDIX-I
Normal Distribution:

$$P.D.F.: f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2]$$

$$C.D.F.: F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-infinity}^{x} \exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2]$$

Parameters: μ = location parameter, σ = scale parameter

Reduced Variate: $Z = \frac{(x-u)}{\sigma}$ (41)

P.D.F.: $f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^{2/2}}$ (42)

$$\text{C.D.F.}: F(z) = -\int_{-infinity}^{z} \frac{1}{\sqrt{2\pi}} e^{-z^{2/2}} dz$$

$$f(x) = \frac{1}{\sigma_y\sqrt{2\pi x}} \exp[-\frac{1}{2}(\frac{\log_e x - \mu}{\sigma_y})^2] \quad (43)$$

Mean of the reduced variate: $\bar{z} = 0$
Standard deviation of reduced variate $\sigma_z = 1$
Coefficient of skewness of the reduced variates = $g_z = 0$

Log Normal Distribution (Two Parameters):

$$\text{P.D.F.}: f(x) = \frac{1}{\sigma_y\sqrt{2\pi x}} \exp[-\frac{1}{2}(\frac{\log_e x - \mu_y^2}{\sigma_y})^2] \quad (44)$$

C.D.F.: $F(x) = \frac{1}{\sigma_y \sqrt{2\pi}} \int_{xsubo}^{x} \exp[-\frac{1}{2}(\frac{\log_e x - \mu_y}{\sigma_y})^2] dx$ (45)

where, $y = \log_e x$

μ = Mean of Y series, σ_y = Standard deviation of Y-series

Parameters: u_y = location parameter, σ_y = scale parameter

Reduced variates, $Z = \frac{\log_e x - \sigma_y}{\sigma_y}$ (46)

P.D.F.: $f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^{2/2}}$ (47)

C.D.F.: $F(z) = \int_0^z - \frac{1}{\sqrt{2\pi}} esup - z^{2/2} dz$ (48)

Mean of the reduced variate: $\bar{z} = 0$

Standard deviation of reduced variate: $\sigma_z = 1$

Coefficient of skewness of the reduced variates = $g_z = 0$

Gumbel Extreme Value (Type-I) Distribution(EV1):

$$P.D.F.\ f(x) = \frac{1}{\alpha} \exp[-(\frac{x-u}{\alpha}) - \exp\{-(\frac{x-u}{\alpha})\}]$$

C.D.F.: $F(x) = e^{-e^{-(\frac{x-u}{\alpha})}}$ (49)

Reduced Variate : $z = \frac{x-u}{\alpha}$ (50)

P.D.F.: f(z) = exp (-z-exp(-z)) (51)

C.D.F. : $F(z) = e^{-e^{-z}}$ (52)

Mean reduced variates Z = 0.5772

Standard deviation of reduced variate $\alpha_z = \frac{\pi}{\sqrt{6}} = 1.2825$

Coefficient of skewness of reduced variate $g_z = 1.14$

Pearson type-III Distribution (PT3):

$$P.D.F.: f(x) = \frac{(x - x_o)^{\gamma-1} e^{-(x-x_0)/\beta}}{\beta^{\gamma} \sqrt{\gamma}}$$

$$C.D.F. F(x) = \int_{x_0}^{x} \frac{(x - xsubo)^{\gamma-1} e^{-(x-x_0)/\beta}}{\beta^{\gamma} \sqrt{\gamma}} dx$$

Parameters: x_o = Location parameter; β = Scale parameter; γ = Shape parameter

Reduced variates, $Z = \frac{x - x_o}{\beta}$ (53)

P.D.F.: $f(z) = \frac{1}{|\beta|\sqrt{\gamma}} (z)^{\gamma-1} e^{-z}$ (54)

C.D.F.: $F(z) = \int_o^z \frac{1}{|\beta|\, sqrtgamma} (z)^{\gamma-1} e^{-z}$ (55)

Mean of the reduced variate: $\bar{z} = \gamma$

St. Deviation of the reduced variate: $\sigma_z = \sqrt{\gamma}$

Coefficient of skewness of the reduced variate: $g_z = 2/\sqrt{\gamma}$

Log Pearson Type-III Distribution (LP3):

$$P.D.F.: f(x) = \frac{(\log_e x - y_o)^{\gamma-1} e^{-(\log_e x - y_o)/\beta}}{|\beta|\sqrt{\gamma}\ x} \text{ ; C.D.F.: } F(x) = \int_{yo}^{x} f(x)dx \quad (56)$$

Parameters: y_o = Location parameter; β = scale parameter; γ = shape parameter

Reduced variate: $Z = \frac{\log_e x - y_o}{\beta}$ (57)

P.D.F.: $f(z) = \frac{1}{|\beta|\gamma} (z)^{\gamma-1} e^{-z}$ (58)

C.D.F.: $F(z) = \int_o^z f(z)dz$ (59)

Mean of reduced variate = $\bar{z} = \gamma$

Standard dev. of reduced variate: $\sigma_z = \sqrt{\gamma}$

Coefficient of skewness of reduced variate: $g_z = 2/\sqrt{\gamma}$

General Extreme Value Distribution:

$$f(x) = \frac{1}{\alpha}\left[1 - k\left(\frac{x-u}{\alpha}\right)\right]^{\frac{1}{k}-1} e^{-\left[1-k\left(\frac{x-u}{\varepsilon}\right)\right]^{\frac{1}{k}}}$$

C.D.F.: $F(x) = \text{Exp } [-(1 - K(\frac{x-u}{\alpha}))^{1/K}]$

Parameters: u = Location parameter, α = Scale parameter, K = Shape Parameter

If K = 0 it leads to EV-I distribution,

K < 0 it leads to EV-II distribution
K > 0 it leads to EV-III distribution

GEV reduced variates, $w = \frac{x - u}{\alpha}$ (60)

Here w = (1_e - kz)/K and Z = EV-I reduced variate

Gamma Distribution:

It is a special case of PT3 distribution.

P.D.F.: $f(x) = \frac{(x)^{\gamma-1} e^{-x/\beta}}{betasup\gamma \sqrt{\gamma}}$ .... (61)

C.D.F.: $F(x) = \int_o^x \frac{(x)^{\gamma-1} e^{-x/\beta}}{\beta^\gamma \sqrt{\gamma}} dx$ (62)

Parameters: β = Scale Parameter

γ = Shape parameter

Gamma Reduced Variate $z = \frac{x}{\beta}$ (63)

P.D.F.: $f(z) = \frac{1}{|\beta| \sqrt{\gamma}} (z)^{\gamma-1} e^{-z}$ (64)

C.D.F.: $F(z) = \int_o^z \frac{1}{|\beta| \, sqrtgamma} (z)^{\gamma-1} e^{-z}$ (65)

Mean of the reduced variate: $\bar{z} = \gamma$

St. Deviation of the reduced variate: $\sigma_z = \sqrt{\gamma}$

Coefficient of skewness of the reduced variate : $g_z = 2/\sqrt{\gamma}$

Exponential Distribution:

P.D.F.: $f(x) = \frac{1}{\beta} e \frac{-(x - x_o)}{\beta}$ (66)

C.D.F.: $F(x) = 1 - e \frac{-(x - x_o)}{\beta}$ (67)

Parameters: x_o = Location Parameter

β = Scale Parameter and Reduced Variates, $z = \frac{x - x_o}{\beta}$ (68)

P.D.F.: $f(z) = e^{-z}$ (69)

C.D.F.: $F(Z) = 1 - e^{-z}$.(70)

Mean of the reduced variate: $\bar{z} = 1$

St. Deviation of the reduced variate $\sigma_z = 1$

Coefficient of skewness of the reduced variate : $g_z = 2$