Training Course

On

# Hydrological Processes in an Ungauged Catchment

[July 25 – 29, 2011]

*CHAPTER-3*

## Flood Frequency Analysis Techniques

By


Rakesh Kumar
Scientist-F


Organized under Hydrology Project - II


National Institute of Hydrology
Roorkee – 247667 (Uttrakhand)

# FLOOD FREQUENCY ANALYSIS TECHNIQUES

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

## 3.0 INTRODUCTION

Hydrological processes are generally chance and time dependent processes. Probabilistic modeling considers only the probability of occurrence of an event with a given magnitude and uses probability theory for decision making. As the probabilistic modeling complete ignores the time dependence of the process it can be used only for design purposes and not for operational purposes.

Probabilistic modeling or frequency analysis is one of the earliest and most frequently used applications of statistics in hydrology. Early applications of frequency analysis were largely in the area of flood flow estimation but today nearly every phase of hydrology is subjected to frequency analysis is to assume the specific probability distribution which the event is likely to follow and to proceed to evaluate the parameters of the distribution using the available data of the events to be modeled. Using the statistically derived sample estimates, probability levels can be assigned to any specific event and the prediction can be made for the required probability event. Though the present lecture deals with the frequency analysis of flood flow only yet the concepts given can be extended for other hydrological variables also.

In fact, information on flood magnitudes and their frequencies is needed for design of hydraulic structures such as dams, spillways, road and railway bridges, culverts, urban drainage systems, flood plain zoning, economic evaluation of flood protection projects etc. Pilgrim and Cordery (1992) mention that estimation of peak flows on small to medium-sized rural drainage basins is probably the most common application of flood estimation as well as being of greatest overall economic importance. These estimates are required for the design of culverts, small to medium-sized bridges, causeways and other drainage works, spillways of farm and other small dams and soil conservation works. For this purpose, statistical flood frequency analysis has been one of the most active areas of research since the last forty to fifty years. Flood frequency analysis is expected to provide solutions to some of the questions such as (i) Which parent distribution the data may follow? (ii) What should be the most suitable parameter estimation technique? (iii) How to account for sampling variability while identifying the distributions? (iv)What should be the suitable measures for selecting the best fit distribution? (v) What criteria one should adopt for testing the regional homogeneity? The procedures of finding better solutions to the aforementioned questions have improved with the efforts made by the various investigators. The scope of frequency analysis would have been widened if the parameters of the distribution could have been related with the physical process governing floods. Such relationships, if established, would have been much useful for studying the effects of non-stationary and man made changes in the physical process on frequency analysis. In spite of many limitations, the statistical flood frequency analysis remains the most important means of quantifying floods in systematic manner.

As such there are essentially two types of models adopted in flood frequency analysis literature: (i) annual flood series (AFS) models and (ii) partial duration series models (PDS). Maximum amount of efforts have been made for modelling of the annual flood series as compared to the partial duration series. In the majority of research projects attention has been confined to the AFS models. The main modelling problem is the selection of the probability distribution for the flood magnitudes coupled with the choice of estimation procedure. A large number of statistical distributions are available in literature. Among these the Normal, Log Normal, Gumbel, General Extreme Value, Pearson Type III, Log Pearson Type III, Generalized Pearson, Logistic, Generalized Logistic and Wakeby distributions have been commonly used in most of the flood frequency studies. For the estimation of the parameters of the various distributions the graphical method, method of least squares, method of moments, method of maximum likelihood, method based on principle of maximum entropy, method of probability weighted moment and method of L-moment are some of the methods which have been most commonly used by many investigators in frequency analysis literature. Once the parameters are estimated accurately for the assumed distribution, goodness of fit procedures then test whether or not the data do indeed fit the assumed distribution with a specified degree of confidence. Various goodness of fit criteria have been adopted by many investigators while identifying the best fit distribution from the various distributions fitted with the historical data.

The broad area of flood frequency analysis has been covered in the light of the following topics:

(i) Definitions
(ii) Assumptions and data requirement
(iii) Plotting positions
(iv) Commonly used distributions in flood frequency analysis
(v) Parameter estimation techniques
(vi) Goodness of fit tests and
(vii) Estimation of T year flood and confidence limits

## 3.1 DEFINITIONS

a.) Peak Annual Discharge: The peak annual discharge is defined as the highest instantaneous volumetric rate of discharge during a year.

b.) Annual flood series: The annual flood series is the sequence of the peak annual discharges for each year of the record.

c.) Design Flood: Design flood is the maximum flood which any structure can safely pass. It is the adopted flood to control the design of a structure.

d.) Recurrence interval or return period: The return period is the time that elapses on an average between two events that equal or exceed a particular level. For example, T year flood will be equaled or exceeded on an average once in T years.

e.) Partial flood series: the partial flood series consists of all recorded floods above a particular threshold regardless of the number of such floods occurring each year.

f.) Mean: Mean is a measure of central tendency. Other measures of central tendency are median and mode. Arithmetic mean is the most commonly used measure of central tendency and is given by

$$\bar{x} = \sum_{i=1}^{N} x_i / N \tag{3.1}$$

where $x_i$ is the ith variate and N is the total number of observations.

g.) Standard Deviation: An unbiased estimate of standard deviation ($S_x$) is given by

$$S_X = (\sum_{i=1}^{N} (x_i - \bar{x})^2 / N - 1)^{0.5} \tag{3.2}$$

Standard deviation is the measure of variability of a data set. The standard deviation divided by the mean is called the coefficient of variation and ($C_v$) is generally used as a regionalization parameter.

h.) Coefficient of skewness ($C_s$) : The coefficient of skewness measure the assymtry of the frequency distribution of the data and an unbiased estimate of the $C_s$ is given by

$$C_s = \frac{N\sum_{i=1}^{N}(x_i - \bar{x})^3}{(N-1)(N-2)S_x^{\ 3}} \tag{3.3}$$

i.) Coefficient of kurtosis ($C_k$) : The coefficient of kurtosis is $C_k$ measures the peakedness or flatness of the frequency distribution near its centre and an unbiased estimate of it is given by

$$C_k = \frac{N^2\sum_{i=1}^{N}(x_i - \bar{x})^4}{(N-1)(N-2)(N-3)S_x^{\ 4}} \tag{3.4}$$

j) Probability paper : A probability paper is a specially designed paper on which ordinate represents the magnitude of the variable and abscissa represent the probability of exceedance or nonexceedance. Proability of exceedance, $P_r(X \geq x)$, probability of non exceedance, $P_r(X \leq x)$ and return period (T) are related as

$$P_r(X > x) = 1 - P_r(X \leq x)$$

$$P_r(X > x) = 1/T$$

Plotting position formulae are used to assign probability of exceedance to a particular event.

## 3.2 ASSUMPTIONS AND DATA REQUIREMENT

### 3.2.1 Assumptions:

The following three assumptions are implicit in frequency analysis.

(i) The data to be analyzed describe random events,
(ii) The natural process of the variable is stationary with respect to time and
(iii) The population parameters can be estimated from the sample data.

### 3.2.2 Data Requirement:

For flood frequency analysis either annual flood series or partial duration flood series may be used. The requirements with regard to data are that

(i) It should be relevant,
(ii) It should be adequate and
(iii) It should be accurate.

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 guide line for the length of data to be used for frequency analysis. Generally a length of 30 – 35 years is considered adequate for flood 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 nonhomogeneous 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.

## 3.3 PLOTTING POSITIONS

In order to see the fit of the distribution, the sample data is plotted on various probability papers. The plotting position formulae are required to assign probability of exceedance or nonexceedance to a particular event. The general plotting position formula is given by

$$p(x \geq x) = \frac{m-a}{N+1-2a} \tag{3.5}$$

Where p(x > x) is the probability of exceedance, m is the rank of the event when arranged in descending order and N is the total number of observations. For the largest value m will naturally be 1 while for the lowest value it will be equal to N. Varies plotting position formulae have been proposed in the literature. Given below are some of the formulae and the values of a in the formula

| Fomula | Value of a |
|---|---|
| Weibull | 0 |
| Blom | 3/8 |
| Gringortan | 0.44 |

Weibull is the most commonly used formula while Blom and Gringortan have been recommended for Normal and Gumbel distributions respectively. NERC (1975) recommends a = 0.4 for plotting position formula to be used for Pearson type III and log Pearson type III distributions.

Probability papers for normal, log normal and Gumbel distributions are given in fig3.1, 3.2 and 3.3 respectively.

## 3.4 COMMONLY USED DISTRIBUTIONS

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. Some of the probability distributions which are commonly used in frequency analysis are explained in brief in subsequent sections.

### 3.4.1 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 unique 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 PDF and CDF of the distribution are given in Appendix.

### 3.4.2 LOG NORMAL DISTRIBUTION

The causative factors for many hydrologic variables act multiplicatively rather 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, then x is said to follow log normal distribution. If the variable x has a lower bound $x_0$, different from zero and the variable $Y = \log(X-x_0)$ follows normal distribution then X is log normally distributed with thee parameters.

### 3.4.3 PEARSON TYPE III DISTRIBUTION

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 distribution are given in appendix.

### 3.4.4 LOG PEARSON TYPE III DISTRIBUTION

If $Y = \log_e(X)$ follows Pearson type III distribution then X is said to follow log Pearson type III distrigution.

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.

### 3.4.5 GUMBEL (EXTREME VALUE TYPE 1) DISTRIBUTION

One of the most commonly used distributions in flood frequency analysis is the double exponential distribution (known as Gumbel distribution or extreme value type1 or Gumbel EV1 distribution). The CDF of EV-1 distribution is defined as

$$F(x) = \exp(-\exp(-(x-u)/\alpha)) \qquad (3.6)$$

Where u and $\alpha$ are the location and scale parameters of the distribution. Using method of moments, u and $\alpha$ are obtained by following equation.

$\bar{x} = u + 0.5772\alpha$ and

$$S^2_X = \pi^2\alpha^2/6 \qquad (3.7)$$

Above equation can be written in the reduced variate form as

$$F(x) = \exp(-\exp)(-y)) \qquad (3.8)$$

where

$y = (x-u)/\alpha$

The reduced variate y can be written in terms of return period, T, also by replacing F(x) by 1-1/T as

$$Y = -\ln(-\ln(1-1/T)) \qquad (3.9)$$

$= -\ln(\ln(T/(T-1)))$

or

$$X_T = u - \alpha \ln.\ln(T/(T-1)) \tag{3.10}$$

For Gumbel distribution coefficient of skewness is equal to 1.139.

### 3.4.6 General Extreme value Distribution (GEV)

Jenkinson (1969) suggested a single equation of following type for the GEV distribution.

$$F(x) = \exp(-(1-k((x-u)/\alpha))^{1/k}), \quad \text{for } k\neq 0 \tag{3.11}$$

$$f(x) = \exp(-\exp(-(x-u)/\alpha)) \qquad \text{for } k = 0$$

In the above equation, u, $\alpha$ and k are location, scale and shape parameters respectively. The shape parameters (k) and coefficient of skewness are interrelated. For EV1 distribution k is 0.0 and coefficient of skewness is equal to 1.139. for EV2 distribution k is –ve and $C_s$ is greater than 1.139 for EV3 distribution k is +ve and $C_s$ is letter than 1.139.

### 3.4.7 WAKEBY DISTRIBUTION

Houghton (1978) found that a significant majority of high quality flood records cannot be modeled adequately by conventional distributions and so advanced the Wakeby distribution for fitting food records. The Wakeby distribution defined implicity in inverse form as

$$x = m + a(1-F^{b}) - c(1-F^{-d}) \tag{3.12}$$

where

F is the nonexceedance probability of x,
a, b, c and d are positive constants and
m is a location parameter. The parameters a and b govern the left hand tail (drought) while parameter m, c and d govern flood flows.

Being a five parameter distribution, the Wakeby distribution can span regions of distribution function space inaccessible to the conventional two or three parameter distributions, besides span regions occupied by the conventional distributions. This implies that the Wakeby distribution can mimic the form of those two parameter distributions whose inverse form exist. Wakeby distribution is currently becoming popular in USA among researchers due to its capability to model both the tail ends of the data separately. Hydrologists and engineers in the past years have occasionally felt the need to go beyond three parameters but it was recognized that the use of moments higher than the third would introduce too much error into the estimation process. The estimation procedure development for the Wakeby distribution circumvents this problem. Main features of the Wakeby distribution over the traditional distributions are

(i) In traditional estimation procedures, the smallest observation can have a substantial effect on the right hand side (large observations) of the distribution. But the left hand side/small observations do not necessarily add information to an estimation of quantile on the right hand side. Since floods are not known to follow any particular distribution, it seems intuitively better to model the left and right hand tails separately.

(ii) None of the traditional distributions have properties to reflect the nature of their left tails accurately. If in reality, the low observations follow the left hand tail of the low skewed distribution and highest observation follow the right hand tail of highly skewed distributions, then none of the conventional three parameter distributions will be able to model it accurately. They lack enough kurtosis for any given skew. Fitting a three parameter curve to a five parameter nature would distort the whole fit including the higher quantiles.

(iii) The separation effect which refer to the differences appear between samples of synthetic stream flow data and natural streamflow data when the standard deviation of skew is plotted versus the mean of skew for regional data, is explained by Wakeby distribution.

(iv) Given the correct choice of parameters, the Wakeby distribution can generate synthetic flows in the pattern of most of the conventional distribution.

## 3.5 PARAMETER ESTIMATION TECHNIQUES

Various parameter estimation techniques which are in current use include

(i) Graphical method
(ii) Least squares method
(iii) Method of moments
(iv) Method of maximum likelihood
(v) Method based on principle of maximum entropy and
(vi) Method based on probability weighted moments.

For further details lecture notes of Workshop on 'Flood Frequency Analysis', NIH (1987-88) may be consulted.

For GEV and Wakeby distributions method based on probability weighted moments has proved to be the most robust method of parameter estimation.

## 3.6 GOODNESS OF FIT TESTS

The validity of a probability distribution function proposed to fit the empirical frequency distribution of a given sample may be tested graphical and analytical methods. Graphical method are usually based on comparing visually the probability density function with the corresponding empirical density function of the sample under consideration and/or the model CDF with the empirical CDF. Often these CDF plots as a straight line. An example of this is the Gumbel paper. If the Gumbel CDF is plotted on Gumbel paper it follows a straight line. If the empirical CDF well approximates a straight line on the Gumbel paper, it is an indication that the Gumbel distribution may be a valid distribution 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 distribution. Some of the commonly used tests are (i) chi-square test, (ii) Kolmogorov-Smirnow test, and (iii) D-index powerful in the sense that the probability of accepting the hypothesis when it is in fact false is very high when these tests are used. In this light D index test is bit better and hence given in subsequent section.

### 3.6.1 D-Index test

The D-index for the comparison of the fit of various distributions in upper tail is given as

$$\text{D index} = (1/\bar{x})\sum_{i=1}^{6} Abs(x_i - \hat{x}_i) \quad (3.13)$$

where $x_i$ and $\hat{x}_i$ are the ith highest observed and computed values for the distribution. The distribution giving the least D-index is considered to be the best fit distribution.

## 3.7 ESTIMATION OF T YEAR FLOOD

T year flood estimated can be obtained either graphically or analytically. Graphically approach is applicable only for normal, log normal and Gumbel EV1 distribution as for other distributions probability papers are not readily available. The main drawback of graphical method is that different engineers will get different estimates fo T year flood. Analytical approach of estimating T year flood for following cases is given in subsequent sections.

(i) Normal distribution (parameter estimation by MOM)
(ii) Log normal distribution (parameter estimation by MOM)
(iii) EV1 distribution (parameter estimation by MOM)
(iv) Pearson Type III (parameter estimation by MOM)
(v) Log Pearson Type III (parameter estimation by MOM)

### 3.7.1 NORMAL (mom)

T year flood $X_T$ is given by

$$X_T = \bar{x} + K_T S_X \quad (3.14)$$

where,

$\bar{x}$ = sample mean
$S_X$ = sample standard deviation
$K_T$ is frequency factor corresponding to probability of exceedance = 1/T and $C_s$ of skewness equal to 0.0. The $K_T$ is the obtained from table of frequency factors (Appendix) adopted from WRC (1981).

### 3.7.2 Log Normal Distribution (MOM)

$$X_T = e^{(\bar{Y} + K_T . S_Y)} \quad (3.15)$$

Where

$\bar{y}$ = mean of log (base e) transformed series
$S_y$ = standard deviation of log transformed series
$K_T$= is frequency factor corresponding to prob. of exceedance equal to 1/T and $C_s$ equal to 0.0 (Appendix)

### 3.7.3 EV1 Distribution (MOM)

$$X_T = u + \alpha . Y_T \quad (3.16)$$

Where

$u = \bar{x} - 0.5772\alpha$

$\alpha = (\sqrt{6}/\pi)\, S_X$
$Y_T$ = reduced variate
$= -\ln(-\ln(1-1/T))$

### 3.7.4 Pearson Type III (MOM)

$$X_T = \bar{x} + K_T\, S_X \tag{3.17}$$

where
$K_T$ = frequency factor corresponding to $C_s$ of original series and prob. of exceedance = 1/T (Appendix).

### 3.7.5 Log Pearson Type III (MOM)

$$X_T = e^{(\bar{Y} + K_T . S_Y)} \tag{3.18}$$

where
$K_T$ = frequency factor corresponding to $C_s$ of log transformed series and prob. of excee. =1/T.

## 3.8 DETERMINATION OF CONFIDENCE ITNERVALS

Hydrologic variables such as annual peak floods or rainfalls do not occur in a set pattern and are mostly random. In modeling these events, the help of frequency analysis is taken such that the estimate of these hydrological variables for a desired return period can be estimated with a reasonable accuracy. The estimates, usually, arrived from a single set of sample data are variable because of randomness associated with these events and the size of the sample used for arriving at the estimates.

Moreover, the sample under consideration is assumed to have resulted from a specific parent population and is random. This results in the fact that there are many equally likely possible samples that can originate from this assumed population. If estimates of the variables for all such samples for the desired return period are plotted against the return period, they seem to follow a normal or t-distribution with its mean as the expected value of the variables at that return period (Fig3.4). This therefore indicates that due to sampling variation there can be many estimates and therefore, should be defined through a continuous run of estimates rather than single or point value of the estimate. This range is defined as confidence interval and can written as

$$\text{Prob}\,(x_{TL} \leq x_T \leq x_{TU}) = 1 - \alpha \tag{3.19}$$

Where $x_{TL}$ and $x_{TU}$ are lower and upper confidence limits of the estimate $x_T$ so that the interval $x_{TL}$ to $x_{TU}$ is the confidence interval and $1-\alpha$ is the confidence level ($\alpha$= significant level). This can also be graphically represented as in fig3.5.

However, the confidence level based on probability values give rise to the limits on either side of curve developed by frequency analysis to indicate the reliability of the estimates as well as the fit.

## 3.9 DEVELOPMENT OF CONFIDENCE BAND

The confidence limits are computed using the following steps :

(i) Choose a statistical distribution for modeling the annual peak flood data at the given station.
(ii) Estimate the parameters
(iii) Compute the quantiles for the desired return period using the estimated parameters.
(iv) Compute the standard error [$S_e(x_T)$] of the estimates (Appendix).
(v) Compute the t-statistics for the desired confidence level (i.e. 1-$\alpha$/2, considering $\alpha$/2%, significance on both sides) and (N-n) degrees of freedom where N=sample size and n = no. of parameters in the distribution selected.
(vi) Compute the upper and lower confidence limits of the quantiles ($x_T$) as

$$x_{TU} = x_T + t_{(N-n)(1-\alpha/2)}.S_e(x_T) \quad (3.20)$$

$$x_{TL} = x_T - t_{(N-n)(1--\alpha/2)}.S_e(x_T) \quad (3.21)$$

(vii) Plot them on either side of the plot of quantiles and join the points on the upper and lower region to given the confidence band.

## 3.10 SHAPE OF CONFIDENCE BAND

The upper and lower confidence limits computed as per pervious section when plotted against various return periods show the minimum difference near the mean values with a diverging trend away from it.

The interval between them for a particular return period increase with the decrease of sample size. This is mainly because of large sampling variance and hence because of large standard errors.

## REFERENCES


1. Haan, C.T. (1977), 'Statistical Methods in Hydrology', Iowa State University Press, Ames. Iowa, USA.
2. Houghton, J.C. (1978), 'The incomplete Mean Estimation Procedure Applied to Flood Frequency Analysis' Water Resources Research, Vol.14, pp.1111-1115.
3. Jenkinson, A.F. (1969), 'Statistics of Extremes in 'Estimation of Maximum Floods', WMO, Technical Note No.98, pp.183-228.
4. Kite, G.W. (1977), 'Frequency and Risk Analysis in Hydrology', Water Resources Publication, Colorado, 80161.
5. NERC (1975), 'Flood Studies Report', Natural Environmental Research council, London.
6. NIH (1987-88), 'Flood Frequency Analysis' workshop lecture notes.
7. Water Resources Council (1981), 'Guidelines for Determining Flood Flow Frequency', Bulletin 17B, Washington D.C.

## STANDARD ERRORS FOR COMMONLY USED DISTRIBUTIONS

**Normal Distribution**

Standard error of quantile estimate $x_T$ is given by

$$SE\,(x_T) = (S_x/\sqrt{N})\,(1 + 0.5\,K_T^2)^{0.5} \qquad (1)$$

**Log Normal Distribution**

The standard error of quantile estimate in log transformed domain is

$$SE\,(y_T\;(S_y/\sqrt{}\,(1+0.5\,K_T^{\;2})0.5 \qquad (2)$$

The average standard error in natural domain is given by

$$SE\,(X_T) = [X_T\{e^{SE\;(Y_T)} - 1\} - X_T\,e^{-SE\;(Y_T)} - 1\}]/2 \qquad (3)$$

Gumbel EV1 Distribution

$$SE(X_T) = (\alpha/\sqrt{N})\,(1.170+0.196Y_T + 1.099Y_T^2)^{\,0.5} \qquad (4)$$

Where $\alpha$ and $Y_T$ scale parameter and reduced variate respectively.

### Probabilistic Modelling – Case Study

**Example**

Compute 500 and 100 years return period floods for river Narmada at Garudeswar assuming that annual flood series (given in Table 1 alongwith statistical parameters) follow:-

(a) log normal distribution (b) log Pearson type III distribution and (c) Gumbell EV-1 distribution. Based on D index test also determine the distribution of the peak floods series.

**Table 1 :** Annual flood series (cumecs) from 1948 to 1970 for the river Narmada at Garudeswar

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 23890 | 26810 | 45630 | 10380 | 13290 | 17100 | 28650 | 29150 | 12910 | 26700 |
| 19700 | 38800 | 21250 | 43360 | 38880 | 14250 | 19560 | 15250 | 13000 | 22670 |
| 58100 | 31170 | 69400 | 18980 | 47980 | 61350 | 27300 | 33750 | 19500 | 22700 |
| 34260 | 38200 | | | | | | | | |

**Statistical Parameters**

| | Original series | Log transformed series |
|---|---|---|
| Mean | 29556.9 | 10.179 |
| Standard deviation | 14864.4 | 0.488 |
| $C_s$ | 1.052 | 0.1 |

**Solution**

Log Normal Distribution

The calculations for 500 and 1000 years flood are shown in table given below :

| Return period (years) | Prob. of exceed. (P=1/T) | Frequency Factor, $K_T$ | $X_T$ in log domain | $X_T$ in orig. domain |
|---|---|---|---|---|
| 500 | .002 | 2.87816 | 11.5835 | 107312.5 |
| 1000 | .001 | 3.09023 | 11.687 | 119014.43 |

The frequency factors have been taken from Appendix II corresponding to P = .002 and .001 and coefficient of skewness equal to zero.

**D Index**

The highest six observations are 69400, 61350, 58100, 47980, 45630 and 43360 cumecs respectively.

The calculation of D index

| Rank | $x_i$ | Pr.=m/N+1 | $K_T$ | $\hat{x}_i$ | Abs($x_i$-$x_i$) |
|---|---|---|---|---|---|
| 1 | 69400 | 1/33 | 1.886 | 66129 | 3271 |
| 2 | 61350 | 2/33 | 1.5678 | 56618 | 4732 |
| 3 | 58100 | 3/33 | 1.3476 | 50850 | 7250 |
| 4 | 47980 | 4/33 | 1.1882 | 47044 | 936 |
| 5 | 45630 | 5/33 | 1.0549 | 44081 | 1549 |
| 6 | 43360 | 6/33 | 0.9216 | 41350 | 2055 |
| | | | | | $\sum$ 9793 |

D index = 19793 / 29556.9 = 0.67

**Log Pearson Type III Distribution**

| Return Period (years) | Prob. of exceed. (P=1/T) | Frequency Factor, $K_T$ | $X_T$ in log domain | $X_T$ in orig. domain |
|---|---|---|---|---|
| 500 | 0.002 | 2.99978 | 11.6428 | 113868.5 |
| 1000 | 0.001 | 3.23322 | 11.7568 | 127618.4 |

The frequency factors halve been taken from Appendix II corresponding to P = 0.002 and 0.001 and coefficient of skewness equal to 0.1.

D-index

| Rank | $X_i$ | Probability P = m/m+1 | $K_T$ | $\hat{x}_i$ | Abs($x_i$- $\hat{x}_i$ ) |
|---|---|---|---|---|---|
| 1 | 69400 | 0.0303 | 1.9348 | 67723 | 1677 |
| 2 | 61350 | 0.606 | 1.6193 | 58059 | 3291 |
| 3 | 58100 | 0.909 | 1.3683 | 51366 | 6734 |
| 4 | 47980 | 0.1212 | 1.1961 | 47226 | 754 |

| | | | | | |
|---|---|---|---|---|---|
| 5 | 45630 | 0.1515 | 1.0576 | 44139 | 1491 |
| 6 | 43360 | 0.1818 | 0.9191 | 41255 | 2105 |
| | | | | | $\sum$ 16052 |

D index = 16052 / 29556.9 = 0.543

**Gumbel EV1 Distribution**

The parameters u and $\alpha$ of EV1 distribution are estimated from the following equations :

$u = \bar{x} - 0.5772\,\alpha$

$\alpha = (\sqrt{6} / \pi)\, S_x$

$\alpha = 11589.725$

$u = 22867.3$

$Y_T = -\ln(-\ln(1-1/T))$

$Y_{500} = -\ln(-\ln(1-1./500.))$

$= 6.2136$

$X_{500} = u + \alpha\, Y_T$

$X_{500} = 22867.3 + 6.2136 \times 11589.725$

$= 94881.21$ cumecs

$Y_{1000} = 6.9072$

$X_{1000} = 102919.81$ cumecs

D-index

| Rank | $X_i$ | Prob. P=m/n+1 | T | $\hat{x}_i$ | $\lvert X_i - \hat{x}_i \rvert$ |
|---|---|---|---|---|---|
| 1 | 69400 | 1/33 | 33 | 63213 | 6187 |
| 2 | 613500 | 2/33 | 16.5 | 55997 | 5353 |
| 3 | 58100 | 3/33 | 11 | 50110 | 7990 |
| 4 | 47980 | 4/33 | 8.15 | 46584 | 1396 |
| 5 | 45630 | 5/33 | 6.6 | 42593 | 3037 |
| 6 | 43360 | 6/33 | 5.5 | 41481 | 1879 |
| | | | | | $\sum$25842 |

D-index = 25842 / 29556.2

D-index is minimum in case of log Pearson type III distribution and hence on the basis of D index test it can be assumed that log Pearson type III distribution fits the data well in the upper tail.