LECTURE—8

SPECIAL PROBLEMS OF FREQUENCY ANALYSIS

OBJECTIVES

The objective of this lecture is to explain various data related problems which one may come across while carrying out flood frequency analysis. The procedure to be adopted in such situations has also been explained in this lecture.

8 1 INTRODUCTION

While carrying out frequency analysis one may encounter following six types of problems in the data.

(i) Broken record
(ii) Incomplete record
(iii) Zero flood years
(iv) Mixed populations
(v) Outliers
(vi) Historical flood data

The definition and statistical treatment of all the above problems are given in subsequent sections.

8.2 BROKEN RECORD

A broken record consists of two or more periods of systematic records separated by unobserved periods. The different record segments are analysed as continuous record with length equal to the sum of segment lengths unless there is some physical change in the watershed between two segments which may make the total record nonhomogeneous.

8.3 INCOMPLETE RECORD

An incomplete record results when some peak flows are missing because these were either too low or too high or the gauge was out of operation during the flood peak. Missing high and low flows data require different treatment.

When one or more high annual peaks during the period of systematic record have not been recorded, there is usually information available from which the peak discharge can be estimated. In most instances the data collecting agency routinely provides such estimates. If not, and such an estimate is made part of the flood frequency analysis, it should be documented and the data collection agency advised.

At some crest gauge sites the bottom of the gauge is not reached in some years. For this situation use of conditional probability adjustment is recommended as described in Appendix VII.

8.4 ZERO FLOOD YEARS RECORD

A record in which some of the years (one or more) are having zero flows is known as zero flood years record.

One solution to tackle this problem is to add a small constant to all of the observations. Another method is to analyse the non zero values and then adjust the relation to the full period of record. This method biases the results as the zero values are essentially ignord. A third and more theoretical sound method would be to use the theorem of total probability. (Haan, 1977)

$$\text{Prob } (X \geqslant x) = \text{Prob } (X \geqslant x \,/\, X = 0).\ \text{Prob } (X = 0) + \text{Prob } (X \geqslant x \,/\, X \neq 0).\ \text{Prob } (X \neq 0). \tag{8.1}$$

Since Prob $(X \geqslant x / X = 0)$ is zero, so

$$\text{Prob } (X \geqslant x) = \text{Prob } (X \neq 0).\ \text{Prob } (X \geqslant x \,/\, X \neq 0) \tag{8.2}$$

In this relationship P $(X \neq 0)$ would be estimated by the fraction of non zero values and prob $(X \geqslant x / X \neq 0)$ would be estimated by a standard analysis of the non zero values with the sample size taken to be equal to the number of non zero values.

Example 8.1

Seventy five years of peak flow data are available from an annual series; 20 of the values are zero, and the remaining 55 values have a mean of 100 Cumecs, standard deviation of 35.1 Cumecs and are log normally distributed, (a) Estimate the probability of a peak exceeding 125 Cumecs and (b) Estimate the magnitude of 25 year peak flow.

Solution :

Prob $(X \neq o) = 55/75 = 0.733$

Z value corresponding to $X = 125$ will be $\dfrac{125-100}{35.1} = 0.712$

Prob $(X \leqslant 125)$ = Area under normal distribution corresponding to $Z = 0.712$ (Appendix-II)

$$= 0.5 + 0.2611 + \frac{0.2642-0.2691}{0.72-0.71} \times (0.712-0.71)$$

$$= 0.5 + 0.2611 + 0.00062$$

$$= 0.5 + 0.26172$$

$$= 0.76172$$

$$\text{Prob } (X \geqslant 125/X \neq 0) = 1\text{-Prob } (X \leqslant 125 / X \neq 0)$$

$$= 1-0.76172$$

$$= 0.23828$$

$$\text{Prob. } (X \geqslant 125) = \text{Prob } (X \geqslant 125 / X \neq 0).\ \text{Prob } (X \neq 0)$$

$$= 0.23828 \times 0.733$$

$$= 0.1746$$

This means, the probability of a peak flow in any year exceeding 125 cumecs is 0.1746.

(b) 25 year flood

Since the statistical parameters of the log transformed series are not given, hence use of theoretical relationships will be made to obtain the parameters of log transformed series

$$\mu_y = \tfrac{1}{2} \ln (\bar{x}^2/(1+\eta^2_x)$$

$$= \tfrac{1}{2} \ln (100^2/(1+(35.1/100)^2))$$

$$= 4.54$$

$$\sigma^2_y = \log_e (1 + \eta^2_x)$$

$$= \log_e (1.123)$$

$$= 0.1161$$

So $\sigma_y = 0.3408$

$$\text{Prob } (X \geqslant x_{25}) = \text{Prob } (X \neq 0)\ .\ \text{Prob } (X \geqslant x_{25} / X \neq o)$$

$$\text{Prob } (X \geqslant x_{25}) = \frac{1}{25} = 0.04$$

$$\text{Prob } (X \neq o) = 0.733$$

So $\text{Prob. } (X \geqslant x_{25} / X \neq 0) = 0.04/0.733 = 0.0545$

k factor corresponding to prob. ef exceedence = 0.0545 is 1.6

$$\text{So } X_{25} = e^{(\bar{y} + k\,\sigma_y)}$$

$$= e^{(4.54 + 1.6 \times 0.3408)}$$

$$= 162 \text{ cumecs}$$

so the magnitude of 25 year peak flood will be 162 cumees.

8.5 MIXED POPULATIONS

Flooding in some watersheds is created by different types of events. For example, flooding in some watershed is created by snowmelt. rainstorms, or by combination of both snowmelt and rain storm. Such a record may not be homogeneous and may require special treatment. Peak flood series with mixed populations results in flood frequency curves with abnormally large skew coefficients reflected by abnormal slope changes when plotted on lognormal probability paper. In some situations the frequency curve of annual events can best be described by computing separate curves for each type of event. The final frequency curve is obtained by combining population frequency curves. The equation for combining two curves is

$$P_c = P_1 + P_2 - P_1P_2 \tag{8.3}$$

where

P_c = exceedance probability of combined populations for a selected magnitude

P_1 = exceedance probability of some selected magnitude for population series 1.

P_2 = exceedance probability of same selected magnitude for population series 2.

If the flood events that are believed to comprise of two or more populations, cannot be identified and separated by an objective and hydrologically meaningful criteria, the record shall be treated as coming from one population.

8.6 OUTLIERS

Outliers are data points which depart significantly from the trend of the remaining data. The retention, modification, deletion of these outliers can significantly affect the statistical parameters computed from the data. All procedures for treating outliers require judgement involving both mathematical and hydrological considerations. The detection and treatment of high and low outliers is outlined in flow chart-1 given in Appendix-XI. Positive coefficient of skewness of log transformed series indicates the presence of high outlier while the negative coefficient of skewness of log transformed series indicates the possibility of low outlier.

In WRC (1981) procedure the following equation is used to detect high outliers

$$X_H = \overline{X} + K_N \, . \, S \tag{8.4}$$

where,

X_H = high outlier threshold in log units

$\overline{X}$ = mean logarithm of systematic peaks (x's) excluding zero flood events, peaks below gauge base and outliers previously detected.

S = standard deviation of x's

K_N = K value from table given in Appendix X, for sample size N.

These K values are equivalent to a one sided test that detects outliers at the 10% level of significance and are based on a normal distribution for detection of single outliers.

If the logarithms of peaks in a sample are greater than X_H then they are considered as high outliers. Flood peaks considered high outliers should be compared with historic flood data and flood information at nearby sites.

If information is available which indicated a high outlier is the maximum in an extended period of time, the outlier is treated as historic flood. If useful historic information is not available to adjust for high outliers, then they should be retained as part of the systematic record. The treatment of all historic flood data and high outliers should be well documented. The procedure and example to adjust the parameters (mean, standard deviation and coefficient of skewness) is given in Appendix VIII and IX respectively.

The equation used for the identification of low outlier is

$$X_L = \overline{X} - K_N \, . \, S \tag{8.5}$$

where

X_L = low outlier threshold in log units and other terms are same as in eq. 8.4.

If an adjustment for historic flood data has previously been made, then the following equation is used to detect low outliers :

$$X_L = M' - K_H \, . \, S'$$

where,

X_L = Low outlier threshold in log units

K_H = K value from table given in Appendix X

M' = Historically adjusted mean logarithm

S' = Historically adjusted standard deviation

If the logarithms of any annual peaks in sample are less than X_L in eq. 8.5 or 8.6, then they are considered as low outliers. Flood peaks considered low outliers are deleted from the record and the conditional probability adjustment given ni Appendix VII is applied.

8.7 HISTORICAL FLOOD DATA

Information which indicates that any flood peaks which occurred before, during, or after systematic record are maximum in an extended period of time should be used in frequency computations. Before such data are used, tha reliability of the data, the peak discharge magnitude, changes in water conditions over the extended period of time and the effects of these on the computed frequency curve must all be evaluated. The adjustment described in Appendix VIII is recommended when historic data are used. All decisions made should be thoroughly documented.

References

1. Haan, C.T. (1977), 'Statistical Methods in Hydrology', The Iowa State University Press, Ames, Iowa.

2. Water Resources Council (1981), 'Guidelines For Determining Flood Flow Frequency' Bulletin 17 B of the Hydrology Committee, Washington, D.C, 20037.