6.10 Analysis Of Stream Gage Data

6.10.1 Introduction

Many gaging stations exist throughout the State where data can be obtained and used for hydrologic studies. If a project is located near one of these gages and the gaging record is of sufficient length in time, a frequency analysis may be made according to the following discussion. The most important aspect of applicable station records is the series of annual peak discharges. It is possible to apply a frequency analysis to that data for the derivation of flood-frequency curves. Such curves can then be used in several different ways.

- If the subject site is at or very near the gaging site and on the same stream and watershed, the discharge for a specific frequency from the flood-frequency curve can be used directly.
- If the facility site is located on the gaged stream but not at the gaging site, transferring the known frequency discharge may be possible. (See Section 6.11)
- If the flood-frequency curve is from one of a group of several gaging stations comprising a hydrologic region, then regional regression relations may be derived. Regional regression relations are usually furnished by established hydrologic agencies and the designer will not be involved in their development. The Connecticut office of the USGS has developed regression equations for rural watersheds in Connecticut. (See Section 6.12.)

Two methods of estimating flood frequency curves from stream gage data are provided in this manual. The first method is a graphical procedure based on the Gumbel method. The second, a statistical method, which makes use of the Log Pearson Type III frequency distribution.

6.10.2 Application

The designer shall use the stream gage analysis findings for design when there is sufficient years of measured or synthesized stream gage record. The most current stream gage data shall be used for frequency analysis. The data should be obtained directly from the Connecticut office of the USGS. The preferred method for analyzing stream gage data is the Log Pearson Type III method with the Gumbel graphical method being used primarily as a check to insure errors are not made — particularly with the frequency estimates of larger floods. Where series discrepancies 20% +/- are encountered in the findings between the two methods, special studies may be required. These special studies shall consist of such things as comparison with regression equations, application of other flood-frequency methods, and the collection and analysis of historical data. Outliers shall be placed into perspective using the procedure found in Water Resources Council Bulletin 17B.

6.10.3 Graphical Method Procedure

In order to develop frequency-discharge information from stream gage data, a record of 10 years or more is required. Gage stations with records which include controlled watershed runoff should be avoided since the natural response to flood events may not have been recorded. The following steps may then be taken to develop a flood-frequency curve representing the data.

Step 1 The peak discharges for each water year are listed in chronological order. A water year extends from October 1 through September 30.

Step 2 After the discharges have been listed, they are then numbered in order of their magnitude; that is, the highest discharge for a particular gaging station is assigned the number 1, then next highest 2 and so on. The numbering system will indicate the relative distribution of floods during a given period of years.

Step 3 The probability of exceedence, also called the plotting position, for each annual peak flow is computed using the formula:

plotting position = M/(N+1) **(6.3)**

Where: N = number of years of record
M = rank of a given flood beginning with 1

The reciprocal of the plotting position, (N + 1)/M, is the recurrence interval in years.

The recurrence interval vs. discharge for each year is plotted on commercially available log-probability paper, with the recurrence interval or plotting position as the x-axis and the magnitude of the associated discharge as the y-axis log scale. The data plots shown on the graph illustrate the frequency distribution of the floods for a given station.

Step 5 If Gumbel paper is used the points should theoretically tend to fall in a straight line. This special paper has been developed so that sample data will plot as a straight line if the data are distributed according to the formula:

$$F(Q) = e^{-x} \tag{6.4}$$

Where:

$$x = e^{\alpha(Q - \beta)} \tag{6.5}$$

$$\alpha = 1.281/S \tag{6.6}$$

$$\beta = Q' - 0.450\ S \tag{6.7}$$

Q = mean flow
Q' = mean peak
S = standard deviation

This paper is not available commercially, but most USGS offices have prepared forms of the paper on which the horizontal scale has been transformed by the double-logarithmic transform of Equation 6.4.

Step 6 The recurrence interval and corresponding discharge is believed reliable up to a point where the points fall appreciably out of line. Values read beyond this point are less reliable. Extrapolation beyond the limits of the plotted points is not recommended.

Step 7 This frequency curve applies only to the point on the stream at which the gage is located. The use of the flood frequency curve is as follows: enter with the desired frequency or RI and read up to the geometric mean line, then move across to the discharge scale for the design discharge. The reverse procedure is used when a discharge is known and information on frequency is desired.

Step 8 Flood frequency curves can also be plotted with stream stage rather than discharge as the ordinate. This is somewhat more convenient when checking high water elevations at a bridge site.

The analysis of gaged data permits an estimate of the peak discharge for the desired return period at a particular site. A best fit line can be drawn through the data points by eye, and the peak flow corresponding to the desired return period could be extracted from the curve. This is a very subjective method and each designer may derive different estimates from the same data set. Experience has shown that statistical frequency distributions may be more representative of naturally occurring floods and can be reliable when used for prediction. Although several different distributions are used for frequency analysis, experience has shown the log-Pearson Type III distribution to be one of the most useful. The log-Pearson III distribution and the process of fitting it to a particular data sample are described in detail in Water Resources Council Bulletin 17B, "Guidelines for Determining Flood Flow Frequency," 1981. The following abbreviated procedure is taken from that publication.

In the course of preparing a frequency analysis for a particular watershed, the designer will undoubtedly encounter situations where further adjustments to the data are necessary. Special handling of outliers, historical data, incomplete data, and zero flow years is covered in detail in Bulletin 17B.

The computer system HYDRAIN provides the Log Pearson III flood frequency analysis. The analysis follows the Bulletin 17B guidelines for the calculation of a log-Pearson frequency curve based on the mean, standard deviation and skewness of the logarithms of the recorded annual peak flows.

6.10.4 Statistical Method Procedure

The log-Pearson Type III distribution is the recommended statistical method. This method is defined by three standard statistical parameters: the mean, standard deviation and coefficient of skew. These parameters are determined from the data sample, which normally consists of the peak annual flows for a period of record. Formulas for the computation of these parameters are given below:

$$\mathbf{Q = (\Sigma X) / N} \qquad \textbf{(mean of logs)} \qquad \textbf{(6.8)}$$

Where: N = number of observations and X is the logarithm of the annual peak.

The standard deviation of logs is:

$$\mathbf{S_L = \{[\Sigma X^2 - (\Sigma X)^2 / N] / [N - 1]\}^{1/2}} \qquad \textbf{(6.9)}$$

The coefficient of skew of logs is:

$$\mathbf{G = [N^2(\Sigma X^3) - 3N(\Sigma X)(\Sigma X^2) + 2(\Sigma X)^3] / [N(N - 1)(N - 2)S_L^3]} \qquad \textbf{(6.10)}$$

Using these three parameters, the magnitude of the flood of the desired frequency can be determined from the equation.

$$\mathbf{\log Q = Q_L + KS_L} \tag{6.11}$$

Where: log Q = the logarithm of the flood magnitude.
Q_L = the mean of the logarithms of the peak annual floods.
K = a frequency factor for a particular return period and coefficient of skew (Values of K for different coefficients of skew and return periods are given in WRC Bulletin 17B).
S_L = the standard deviation of the logarithms of the peak annual flood.

If a flood frequency curve is necessary, then by computing several values of Q for different return periods, the log-Pearson fit to the data can be plotted on standard log probability paper. If the skew of the sample data happens to be equal to zero, the plot of the log-Pearson fit to the data will be a straight line. If the skew is negative the plot will be a curve with a downward concavity. If the skew is positive, the plot will be a curve with upward concavity.

6.10.5 Skew

There are two alternative methods for determining the value of the skew coefficient to be used in calculating the log-Pearson curve fit. The value of skew that is calculated directly from the gage data using the above formula is called the station skew. This value may not be a true representation of the actual skew of the data if the period of record is short or if there are extreme events in the period of record. WRC Bulletin 17B contains a map of generalized skew coefficients of the logarithms of annual maximum streamflows throughout the United States and average skew coefficients by one degree quadrangles over most of the country.

Often, the station skew and the generalized skew can be combined to provide a better estimate for a given sample of flood data. Bulletin 17B outlines a procedure for combining the station skew and the generalized skew to provide a weighted skew.