CHAPTER III

THEORETICAL CONSIDERATION

3.1 Basic Approach of Hydrological Data

Most of the raw data of hydrological statistics supplied by network is in the form of time series. Since the data necessarily cover only a certain period of observation, such data represent only an observed sample of the entire population of the continuous process. While the time or space element of hydrological phenomena may be the same, each phenomenon is characterized by a certain value which varies in time and space. This characterization is called variable and its particular value is a variate. Thus, discharges are variables and flood flows are variates. The variable may be continuous or discrete. The discharge is continuous where as the different size of rocks contained in a bedload sample are discrete.

In hydrology, the value of one variate is usually independent of others and the variable in the equation is called random variable. The random character of hydrological variable is most important for justifying the use of statistical method for their processing. All the time series may be characterized by statistical parameters as follow:

1.) The arithmetic mean is computed as

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{i=n} x_i \qquad \text{--------- (1)}$$

where x_i = variate,

n = total number of variates in the series.

2.) The standard deviation is the most used parameter which indicates the dispersion of variates in the series.

$$S_x = \sqrt{\frac{\sum_{i=1}^{i=n} (x_i - \bar{x})^2}{n}} \qquad \text{-------- (2)}$$

for the very long series or

$$S_x = \sqrt{\frac{\sum_{i=1}^{i=n} (x_i - \bar{x})^2}{n - 1}} \qquad \text{-------- (3)}$$

for the short series.

3.) The reduced variate is the ratio of a variate to the arithmetic mean.

$$b_i = \frac{x_i}{\bar{x}} \qquad \text{-------- (4)}$$

4.) The amplitude is the difference between the maximum value and the minimum value of the variable in a series.

$$a' = |x_{max} - x_{min}| \qquad \text{-------- (5)}$$

A further characteristic of a series is the frequency distribution of the values of the variable and is obtained by frequency analysis of the series. In the computation of hydrological characteristics of a basin, the frequency method is to adapt series of observations of one phenomenon to another phenomenon of the same catchment. An objective mathematical method of relating the series of observation is therefore needed.

3.2 Method of Flood Flow Estimation

The magnitude of flood flow at T years return period is estimated by using frequency analysis of the annual peak discharges. Various methods may be used such as Plotting-position Formula, Gumbel's Formula, Pearson Type III Distribution, etc. In this study, Gumbel's Formula will be used and his theoretical concept is submitted here as shown below.

3.2.1 Gumbel's Formula

If X_1, X_2, X_3, ..., X_n are the extreme values observed in n samples of equal size N and if X is an unlimited, exponentially distributed variable, then the theory of extreme values states that, as n and N approach infinity, the cumulative probability P that any of the n extremes will be less than X approaches the expression

$$P = e^{-e^{-b}} \qquad \text{---------- (6)}$$

where e = the base of Napierian logarithms

b = the reduced variate, and is given by

$$b = a(X - X_f) \qquad \text{---------- (7)}$$

a = the dispersion parameter

X_f = the mode of the distribution

For an infinitly large sample, it can be shown by the theory of extreme values that X_f and a are functions of the arithmetic mean $\bar{X}$ and the standard deviation S_x

$$X_f = \bar{X} - 0.45005\, S_x \qquad \text{-------- (8)}$$

$$a = \frac{1.28255}{S_x} \qquad \text{-------- (9)}$$

where $\bar{X}$ and S_x can be computed by eq.(1), (2) or (3)

Eq. (6) is an expression of probability of nonoccurrence. The return period in years can be computed from

$$T = \frac{1}{1 - P} \qquad \text{-------- (10)}$$

Combining eq. (7), (8) and (9), the result is

$$b = \frac{1.28255}{S_x}(X - \bar{X} + 0.45005\, S_x) \qquad \text{---- (11)}$$

Eq. (6) and (10) yield

$$b = -\ln(-\ln(1 - \tfrac{1}{T})) \qquad \text{-------- (12)}$$

Equating the value of b from eq. (8) and (9) lead to

$$X = \bar{X} - 0.45005\, S_x - \frac{S_x}{1.28255}\ln(-\ln(1 - \tfrac{1}{T})) \qquad \text{-- (13)}$$

In this study X_1, X_2, X_3, ..., X_n are the annual peak discharges (maximum flow for each water year, N = 1) and X is the annual flood at T years return period. Eq.(13) can also be uesd to calculate the monthly flood by replacing the annual peak discharges with the monthly peak discharges.

3.3 Relationship between Flood Flow and Basin Characteristics

In the study of flood, the annual flood at T years return period is generally assumed to be a function of basin characteristics such as basin area, shape number, drainage density and slope of the main stream, etc.

$$Q_T = f(A, S_n, D, S, \ldots..)$$

where Q_T = annual flood at T years return period

A = basin area

S_n = shape number

D = drainage density

S = slope of the main stream

The above function can be shown in the mathematical form as the following exponential equations :-

$$Q_T = K A^{n_1} \quad \text{------} \quad (14)$$

$$Q_T = K A^{n_1} S_n^{n_2} \quad \text{------} \quad (15)$$

$$Q_T = K A^{n_1} D^{n_2} \quad \text{------} \quad (16)$$

$$Q_T = K A^{n_1} S^{n_2} \quad \text{------} \quad (17)$$

$$Q_T = K A^{n_1} S_n^{n_2} D^{n_3} \quad \text{------} \quad (18)$$

$$Q_T = K A^{n_1} D^{n_2} S^{n_3} \quad \text{------} \quad (19)$$

$$Q_T = K\,A^{n_1}\,S^{n_2}\,S_n^{n_3} \quad \text{-------- (20)}$$

$$Q_T = K\,A^{n_1}\,S_n^{n_2}\,D^{n_3}\,S^{n_4} \quad \text{-------- (21)}$$

where n_1, n_2, n_3, n_4 = constants derived by fitting the multiple regression line to the available data and are called "multiple regression coefficient".

K = constant for a particular year of return period

By logarithmic transformation of all variables, the general relationships expressed by the above equations can be transformed to a multiple linear regression of the type represent by eq. (22). And the computation of the constants K, n_1, n_2, n_3 and n_4 for any return period can then be made.

3.4 Multiple Linear Regression and Correlation Analysis

If there are m variables to correlate, including one dependent and m-1 externally independent, the general equation for multiple linear regression is

$$X_1 = B_1 + B_2X_2 + B_3X_3 + \ldots\ldots + B_mX_m \quad \text{---- (22)}$$

where B_1 = the intercept

B_j = the multiple regression coefficient of the dependent variable X_1 on the independent variable X_j with all other variables kept constant (j takes from 2 to m).

Applying the least-squares method for the sum of residuals

$$\Delta_1 = X_1' - B_1 - B_2X_2 - B_3X_3 - \ldots\ldots - B_mX_m$$

which m partial differential equations in B_1, B_2, B_3, ..., B_m give m linear equations :-

$$\left.\begin{array}{l}
B_2\Sigma(\Delta X_2)^2 + B_3\Sigma(\Delta X_2\Delta X_3) + \ldots + B_m\Sigma(\Delta X_2\Delta X_m) = \Sigma(\Delta X_1\Delta X_2) \\
B_2\Sigma(\Delta X_2\Delta X_3) + B_3\Sigma(\Delta X_3)^2 + \ldots + B_m\Sigma(\Delta X_3\Delta X_m) = \Sigma(\Delta X_1\Delta X_3) \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
B_2\Sigma(\Delta X_2\Delta X_m) + B_3\Sigma(\Delta X_3\Delta X_m) + \ldots + B_m\Sigma(\Delta X_m)^2 = \Sigma(\Delta X_1\Delta X_m) \\
B_1 = \bar{X}_1 - B_2\bar{X}_2 - B_3\bar{X}_3 - \ldots\ldots\ldots - B_m\bar{X}_m
\end{array}\right\} \quad (23)$$

where $\bar{X}_i = \Sigma X_i/N$

$$\Delta X_i = \Sigma X_i - \bar{X}_i$$

$$\Sigma(\Delta X_i)^2 = \Sigma X_i^2 - N(\bar{X}_i)^2$$

$$\Sigma(\Delta X_i\Delta X_j) = \Sigma X_iX_j - N\bar{X}_i\bar{X}_j$$

with i taken from 1 to m and N is the sample size.

When there are a small number of variables, either ungrouped or grouped data may be used for the analysis. For more than four variables, the use of multivariate distributions in the form of grouped data will become cumbersome, because of the difficulties in obtaining a suitable representation of the multivariate distributions.

The degree of correlation of a dependent variable X_1 to many externally independent variables in a multi-linear association is measured by a multiple correlation coefficient. It measures the combined influence of all externally independent variables in terms of the difference between the actual and the estimated values of the dependent variable. The coefficient is expressed as

$$R_1 = \frac{s_{e_1}}{s_1}$$

$$= 1 - \frac{S_1^2}{s_1^2}$$

$$= \frac{B_2 \Sigma \Delta X_1 \Delta X_2 + B_3 \Sigma \Delta X_1 \Delta X_3 + \dots + B_m \Sigma \Delta X_1 \Delta X_m}{\Sigma (\Delta X_1)^2} \quad \text{----(24)}$$

where s_{e_1} = the standard deviation of estimated values of X_1

s_e = the standard deviation of actual values of X_1

S_1 = the standard deviation of residuals

The multiple correlation coefficient varies between -1 and +1 the relationship may be ascertained on the basis of the values of the coefficient as follow :

$R_1 = 1$	direct functional dependence,
$0.6 < R_1 < 1$	good direct correlation,
$0 < R_1 < 0.6$	insufficient direct correlation,
$R_1 = 0$	no correlation,
$-0.6 < R_1 < 0$	insufficient reciprocal correlation,
$-1 < R_1 < -0.6$	good reciprocal correlation,
$R_1 = -1$	reciprocal functional dependent.

3.5 Basin Characteristics

There are various basin characteristics which have been defined in many different publications in the past. Some of them to be used in this study are reviewed here.

3.5.1 Basin area

The area A_u of a basin of a given order u is defined as the total area projected upon a horizontal plane, contributing overland flow to the channel segment of the given order and including all tributaries of lower order.

In the general case, the total area A_u of a basin of the order u may be written as

$$A_u = (\Sigma A_1 + \Sigma A_2 + \dots\dots\dots + \Sigma A_{u-1}) + (\Sigma A_{01} + \Sigma A_{02} + \dots\dots\dots + \Sigma A_{0u}) \qquad \text{-----} (25)$$

where A_{Ou} is the area of an interbasin area contributing to a u-order segment.

The area of basin of a given order can be measured by polar plannimeter from a map on which the perimeters have been outlined for each order.

3.5.2 Shape number

The shape of a basin can be expressed in terms of various easily defined and measured factors. One measure of basin shape is given by the quotient of the basin area divided by the length of the main stream squared and is called "shape number".

$$S_n = \frac{A}{L^2} \qquad \text{------ (26)}$$

where S_n = shape number

L = length of the main stream

A = basin area

3.5.3 Drainage density

Drainage density is defined as the ratio of the sum of the length of all stream of any river system to the total area drained by the system.

$$D = \frac{\Sigma L}{A} \qquad \text{------- (27)}$$

where D = drainage density

A = basin area

ΣL = sum of the length of all stream in the river system

Drainage density may be thought of as an expression of the closeness of spacing of channels. It is an important indicator of the linear scale of land-form elements in stream-erode topography.

3.5.4 Slope of the main stream

The slope of the main stream is probably the most significant physical characteristic after the basin area. It can be determined by dividing twice the mean height of the stream profile above the point of interest by the length of the main stream. The mean height of the profile is obtained by integration of the elevation-distance curve (stream profile) and dividing by the length of the stream. The integration can be accomplished by "plannimetering" the area below the elevation-distance curve. This slope is also given by the slope of the line drawn through the origin of the stream-profile curve such that the area under it equals the area under the profile curve.

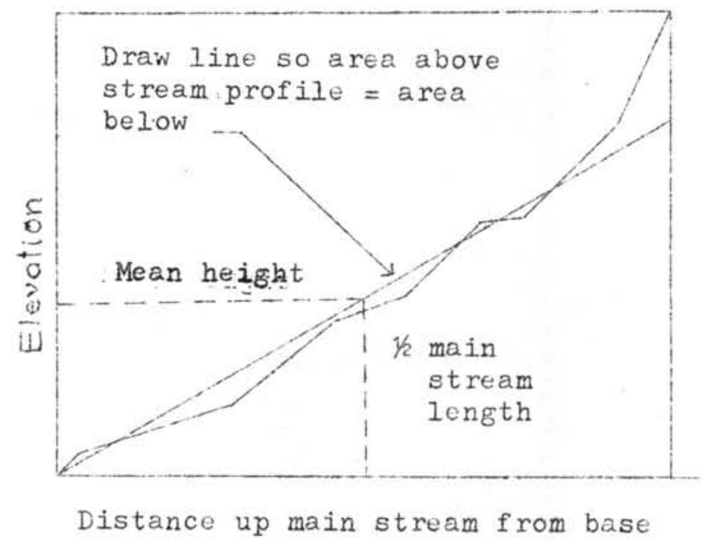

Fig.5 Slope of the main stream

3.5.5 Length of stream

The most commonly used length of stream are L and L_c

1.) L which is called " length of the main stream ", is the longest length measured along the main stream from the point of interest to the drainage-basin divide. This length is normally measured on the topographical map, which along with the projection of the main stream to the basin divide, may cause some minor discrepancies in the length when measured by different individuals. However such discrepancies are generally insignificant in their effect on hydrologic computations.

2.) L_c has been previously defined as the length measured up the stream channel from the base of the drainage area to a point corresponding to the centroid of the area above the area-distance curve. Later, some writers have added physical definitions of L_c which are not rigidly correct while others leave the physical interpretation of L_c purposely vague. Therefore L_c is defined as the distance from the outlet to a point on the stream which is nearest to the centroid of the basin.

Fig. 6 shows a basin area. Axis X-X is drawn perpendicular to the major axis of the area through the base of the area. The true area-centroid is then located a distance $\bar{y}$ above this axis where

$$y = \frac{\Sum y_1 \triangle A}{A} \qquad \text{......... (28)}$$

in which A is an elemental area, y_1 is its distance from the axis X-X, and A is the total area. The term L_c can be expressed by the similar formula

$$L_c = \frac{\Sigma y_2 \triangle A}{A} \qquad \text{......... (29)}$$

where y_2 is the distance measured along the stream to the point on the stream at which overland runoff resulting from rainfall on $\triangle A$ first reaches the stream.

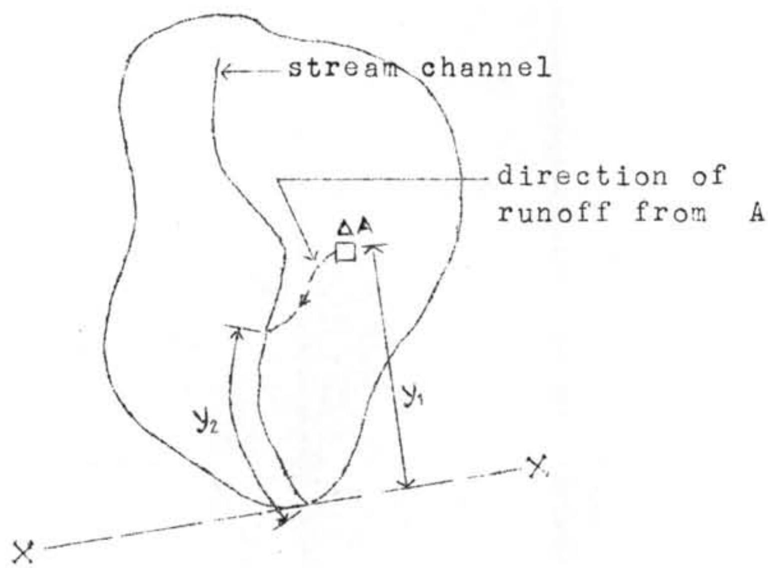

Fig. 6 Hypothetical L_c