APPLIED HYDROLOGY AND SEDIMENTOLOGY FOR DISTURBED AREAS

B.J. Barfield
Professor

and

R.C. Warner
Assistant Professor

Department of Agricultural Engineering
University of Kentucky
Lexington, Kentucky

and

C.T. Haan
Professor and Head

Department of Agricultural Engineering
Oklahoma State University
Stillwater, Oklahoma

Surface detention represents the buildup of water of shallow depth required to support the overland flow process. Surface storage represents water held in depressions or other areas that does not enter the runoff process. Once surface storage elements are filled, they then contribute to the flow processes. The infiltration process is in itself a complex function of time, location and antecedent conditions.

Soil water storage enters the storm water picture only through its affect on infiltration. Infiltration is a decreasing function of soil water content. Dry soils have higher infiltration rates than the same soil when wet. In some cases the soil water storage may be filled resulting in essentially zero infiltration.

It must be kept in mind that this chapter is devoted to hydrology of storm water runoff – not hydrology in general. The difference in these two is that the former deals with the flow and storage processes during and immediately following major precipitation events, while the latter deals with these processes for all time. Only if one is considering the continuous simulation of streamflow is it necessary to treat hydrology in more detail.

Figure 2.4 can be used as a basis for a general description of the storm water runoff process. The objective is to be able to estimate flow resulting from a precipitation event. When precipitation first strikes the surface of the watershed, it becomes either surface storage or surface detention. If a pervious surface is being considered and the infiltration rate exceeds the precipitation rate, then the amount of water in surface storage and surface detention will decrease. These two storages naturally have lower limits of zero at which time the infiltration rate will equal the precipitation rate until the precipitation rate exceeds the rate at which water can infiltrate, and then the two storages will again start to build up.

Surface storage represents water stored on the surface of a watershed which does not become surface runoff. Generally one considers that surface storage must be filled before runoff can start; however, there can be surface runoff from parts of the basin having its surface storage satisfied.

Surface detention is the depth of water attained in the overland

flow process or in the nonchannelized flow process. If we consider a sloping plane surface, we can envision the dynamic nature of surface detention storage. As a rain starts, detention storage builds and runoff starts. The runoff rate will not equal the rainfall rate because some of the water is going into detention storage. If the rainfall rate decreases, surface detention will decrease with the decrease going into runoff. Thus surface detention tends to dampen out rapid changes in rainfall rates and produce slower changes in runoff rates. Of course, the flow velocity of the runoff depends on the flow depth. As the flow depth increases, the flow velocity increases.

Because of surface detention storage and the relationship between flow velocity and flow depth, the runoff hydrograph, even from a smooth impervious surface, is not simply a translation based on area and a travel time of the rainfall pattern. Surface detention and overland flow depth are a function of the roughness of the flow surface and its slope.

Before discussing in detail those elements of the hydrologic cycle important in storm water management, a review will be made of some concepts associated with the stochastic nature of hydrologic events. This will be followed by sections on precipitation, abstractions from precipitation and the runoff process including a detailed treatment of runoff hydrographs and peak flow estimation.

Frequency Analysis

INTRODUCTION

In any discussion of hydrology one constantly hears such terms as the 100-year flood or the 50-year rainfall. Many times these terms are used rather loosely, and rarely are they understood by the layman. Many times the person that is using these terms does not fully appreciate their meaning, the implications associated with them, the difficulty of estimating the magnitude of events associated with the terms and the uncertainty or variability of an estimate for the magnitude of an event associated with the terms.

A generalized notation will be used to denote the events of interest. T-year event denotes an event with a return period of T years (return period is yet to be defined). Q_T will denote the magnitude or peak discharge of a T-year flood. As will be seen, Q_T will never be known with certainty. One must always deal with an estimate for Q_T.

RETURN PERIOD AND PROBABILITY

It is well known that maximum observed streamflow (the peak flow) observed on any stream over a period of one year varies from year to year in an apparently random fashion. This randomness has led to the use of probability and statistics in selecting the hydraulic capacity of storm water facilities. Reference should be made to Haan (1977) for a more complete treatment of this topic. The following is a generalized treatment of hydrologic frequency analysis.

A T-year event is defined as an event of such magnitude that over a long period of time (much, much longer than T years), the average time between events having a magnitude equal to or greater than the T-year event is T years. This being the case, the expected number of occurrences of a T-year event in an N year period would be N/T. For example, one would expect 5 occurrences of a 20-year event in a period of 100 years. This is another way of saying that on the average one expects a T-year event to occur once every T years. It is to be emphasized that there is no regularity associated with a T-year event. It is not to be implied that a T-year event occurs once every T years nor should it be taken that in any T-year period there will always be 1 and only 1 occurrence of a T-year event. In fact, later we will show that there is a chance that in any T-year period, a T-year event can occur 0, 1, 2, ..., T times. Further we will show how to calculate the probabilities of these various possibilities.

The return period of a T-year event as defined above is T years. Often the actual time between occurrences of a T-year event is called the recurrence interval. Thus, the average value of recurrence interval is equal to the return period. Most discussions of return period and recurrence interval assume the two terms are synonymous. Thus, in most instances, when one uses the term recurrence interval, the average recurrence interval is meant.

Since the average time between occurrences of a T-year event is T years, the probability of a T-year event in any given year is 1/T. Thus we have the relationship

$$p_T = 1/T \tag{2.1}$$

where T is the return period associated with an event Q_T and p_T is the probability of Q_T in any given year. Probability is expressed as a number between 0 and 1 inclusively. A probability of 0 means the event cannot happen while a probability of 1 means the event will certainly happen. Sometimes probability is expressed as a percent chance in which case the true probability is multiplied by 100.

So far we have made several assumptions that must be emphasized. The assumptions involve the variable Q, the peak flow in any year. First, we have assumed that the peak flows from year-to-year are independent of each other. This means that the magnitude of a peak in any year is unaffected by the magnitude of a peak in any other year. Secondly, we have assumed that the statistical properties of the peak flows are not changing with time. This means that there are no changes going on within the watershed that results in changes in the peak flow characteristics of the watershed. It further means that the watershed characteristics have remained constant over whatever period of time any data we are using was generated. In the language of statistics, we are assuming the data are from a stationary time series.

Under these assumptions, the occurrence of a T-year event is a random process meeting the requirements of a particular stochastic process known as a Bernoulli process. The probability of Q_T being equaled or exceeded in any year is p for all time and is unaffected by any prior history of occurrence of Q_T. Let us now denote any event equaling or exceeding Q_T as Q_T^*. We do not know the actual magnitude of Q_T^*, we only know that it equals or exceeds Q_T($Q_T^* \geqslant Q_T$). Q_T^* is a Bernoulli random variable. The probability of k occurrences of Q_T^* in n years can be evaluated from the binomial distribution

$$f(k;p_T,n) = \frac{n!}{(n-k)!k!}\,(p_T)^k\,(1-p_T)^{n-k} \tag{2.2}$$

where $f(k;p_T,n)$ is the probability of k occurrences of Q_T^* in n years if the probability of Q_T^* in any single year is p_T[1]. For example, the

1. $n! = n(n-1)(n-2)\ldots(3)(2)(1);\ 0! = 1$

probability of 2 occurrences of a 20-year event in 30 years is

$$f(2;0.05,30) = \frac{30!}{28!2!}(0.05)^2(0.95)^{28} = 0.26$$

In a large number of 30-year records, we would expect 26% of the records to contain exactly 2 peaks that equal or exceed Q_{20}. The other 75% of the 30-year records would contain 0, 1, 3, 4, ..., or 30 peaks that equal or exceed Q_{20}. The probabilities of these later number of exceedances can be evaluated from equation 2.2 also. If this is done, the summation of the probabilities of 0, 1, 2, 3, ..., 30 peaks in 30 years equal to or greater than Q_{20} must equal 1.00 since all possibilities have been exhausted.

Equation 2.2 can be used to calculate the probability that a T-year event will be equaled or exceeded at least once in an n-year period by noting that 'at least once' means one or more. The probability of one or more exceedances plus the probability of no exceedances must equal 1.00. Therefore the probability of at least one exceedance is given by

$$1 - f(0;p_T,n) = 1 - \frac{n!}{0!n!}p_T^{\,0}(1-p_T)^n$$

Since $p_T = 1/T$ and $0! = 1$, this relationship reduces to

$$f(p_T,n) = 1 - (1 - 1/T)^n \qquad (2.3)$$

where $f(p_T,n)$ is the probability that a T-year event will be equaled or exceeded at least once in an n-year period. If n is set equal to T in equation 2.3, it can be shown that for large T, $f(p_T,T)$ approaches the constant 0.632. For T = 10, $f(p_T,T) = f(0.1, 10) = 0.65$. What this means is that if a structure having a design life of T-years is designed on the basis of a T-year event, the probability is about 0.63 that the design capacity will be exceeded at least once during the design life.

By specifying the acceptable probability of the design capacity being exceeded during the design life of the structure, equation 2.3 can be used to calculate the required design return period. For example, if one wants to be 90 percent sure of not exceeding the design capacity of a structure in a 25-year period, $f(p_T,25)$ would be 1 - 0.90 = 0.10. Thus from equation 2.3

$$0.10 = 1 - (1 - 1/T)^{25}$$

or T = 238 years. To be 90 percent sure of not exceeding the design capacity in a 25-year period, the design capacity must be based on an event with a return period of 238 years. In this case the acceptable risk was 10 percent, the degree of confidence was 90 percent, the design life was 25 years and the required design return period was 238 years. Calculations like this can be carried out for various design lifes, design return periods and acceptable risks. Figure 2.5 is based on such calculations and can be used to quickly determine the required design return period based on the design life and acceptable risk or probability of having the design capacity exceeded.

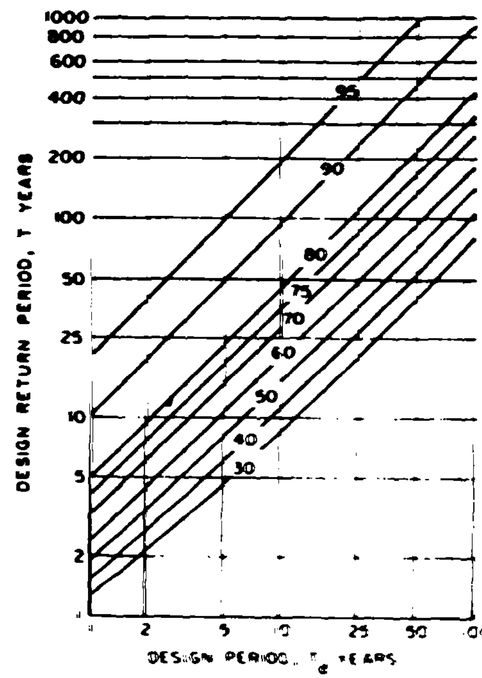

Figure 2.5. Design return period required as a function of design life to be given a percent confident (curve parameter) that the design condition is not exceeded.

In these discussions it should be kept in mind that a high risk of having the design capacity exceeded may be acceptable since what is meant by exceeded is failure of the structure to handle the re-sulting flow in the manner the structure was designed to operate. Failure in this sense does not necessarily mean that the structure will be destroyed. For example, the failure of a road culvert to pass a peak flow may result in only minor flooding of a roadway or adja-cent area and may be acceptable on a fairly frequent basis. On the other hand, failure of a storm water detention basin may result in overtopping of the structure with considerable damage to property and high risk of loss of life downstream. Thus the selection of the acceptable risk and design return period depend on the consequences of the design capacity being exceeded. Building the structure large enough to protect against extremely rare events is quite expensive, while allowing the design capacity to be exceeded on a frequent basis may result in an accumulation of considerable economic loss. Thus the selection of the proper design return period is a problem in economic optimization.

Many governmental units have regulations governing the design period to be used. Often these return periods are based on the size of the structure and the consequences of the structural hydraulic capacity being exceeded. For example, in rural areas road culverts might be based on a 10-year return period. Minor structures in urban areas might be based on the 25-year event and major structures and flood plain delineations might be based on the 100-year event. For example, Table 2.1 shows the design return period specified by Federal Regulation for surface mines as specified in the December 1977 Federal Register.

Table 2.1 Design Return Periods for Certain Facilities Connected with Surface Mines

Item	Return Period
Water Quality Effluent Standards	10-year, 24-hour rain
Settling Ponds	
Volume of Runoff	10-year, 24-hour rain
Spillways (small ponds)	25-year rain
Spillways (large ponds)	100-year, 6-hour rain
Roads	
Out of Flood Plain	100-year
Water Control Structures	10-year

FREQUENCY DETERMINATIONS

Assigning a flood magnitude to a given return period requires knowledge of the flood flow characteristics of the basin of concern. The approach that is used to determine this relationship depends largely on the type, quantity and quality of hydrologic data that is available, and on the importance of the determination. If a minor culvert or channel is to be designed, one cannot justify a time con-suming expensive flood-frequency analysis. On the other hand, if a major component of a drainage system is under construction, the best possible flow estimates are desired.

In this treatment five cases or situations a designer might be faced with are considered:

Case I	- A reasonably long record of streamflow is available at or near the point of interest on the stream of interest.
Case II	- A reasonably long record of streamflow is available on the stream of interest, but at a point somewhat removed from the location of interest.
Case III	- A short streamflow record is available on the stream of interest.
Case IV	- No records are available on the stream of interest, but records are available on nearby streams.
Case V	- No streamflow records are available in the vicinity.

The cases are listed in the order they will be considered. They are also listed in the order of increasing difficulty. Unfortunately, they are listed in the inverse order of their frequency of occurrence. That is, the designer is more likely to be faced with Case V than with Case I, especially for small watersheds. In spite of this, we will devote a major part of our attention to the treatment of Case I. The reason for this is that it is essential that the Case I procedures and their limitation be understood before one can appreciate the problems associated with any of the other cases. The Case I analysis is basic to any

CASE I - FLOOD FREQUENCY DETERMINATION

If one is extremely fortunate, a relatively long record of peak flows on the stream at the point where an estimate for a flood peak of a given frequency is desired may be available. Such a listing might appear as in Table 2.2 for the Middle Fork of Beargrass Creek at Cannons Lane in Louisville, Kentucky.

Table 2.2 Peak Discharge (cfs) Middle Fork Beargrass Creek, Cannons Lane, Louisville, Kentucky.

Year	Peak Flow	Year	Peak Flow	Year	Peak Flow
1945	1810	1956	1060	1966	874
1946	791	1957	1490	1967	712
1947	839	1958	884	1968	1450
1948	1750	1959	1320	1969	707
1949	898	1960	3300	1970	5200
1950	2120	1961	2400	1971	2150
1951	1220	1962	976	1972	1170
1952	1290	1963	918	1973	2080
1953	768	1964	3920	1974	1250
1954	1570	1965	1150	1975	2270
1955	1240				

Any collection of data such as contained in Table 2.2 represents a sample of data from a population. The population in this case would be the maximum annual flood peak for all time, both past and future. The data of Table 2.2 represent a sample from this population. Quantities descriptive of a population are known as parameters. Population parameters are never known in a flood frequency study and must be estimated from the sample of data. Estimates of population parameters are known as sample statistics. Some parameters of interest are the mean, μ_x; the standard deviation, σ_x; the coefficient of variation, C_v; and the skewness, Γ. Sample estimates for μ_x, σ_x, C_v and Γ are given in $\overline{X}$, S_x, C_v and C_s, respectively, and calculated from the equations

$$\overline{X} = \Sigma X_i / n \qquad (2.4)$$

$$S_x = \sqrt{(\Sigma X_i^2 - n\overline{X}^2)/(n-1)} \qquad (2.5)$$

$$\hat{C}_v = S_x / \overline{X} \qquad (2.6)$$

$$C_s = n\Sigma(X_i - \overline{X})^3 / \{(n-1)(n-2)S_x^3\}$$

$$= \frac{n^2 \Sigma X_i^3 - 3n\Sigma X_i \Sigma X_i^2 + 2(\Sigma X_i)^3}{n(n-1)(n-2)S_x^3} \qquad (2.7)$$

where X_i represents the i^{th} data value, n is the sample size and all sumations are from 1 to n. Applying these equations to the Beargrass Creek data results in $\overline{X}$ = 1599 cfs, S_x = 1006 cfs, C_v = 0.619 and C_s = 2.13.

The mean is simply a measure of the central location of a group of data. The standard deviation is a measure of the spread of the data. The larger the standard deviation, the greater the spread in the data. The square of the standard deviation is known as the variance. The units on the standard deviation are the same as the units on the raw data. A dimensionless measure of the spread of a set of data is the coefficient of variation. A compact data set will have a smaller coefficient of variation than will a wide ranging set of data.

The skewness is a measure of the symmetry of a distribution. The normal distribution has a skewness of zero. If the data tends to spread, or tail, to the right more than it does to the left with respect to its mean, the data is positively skewed and C_s will be positive. Data tailing to the left more than to the right is negatively skewed, and C_s will be negative.

If data such as contained in Table 2.2 meet certain assumptions, we can consider them to be independent random variables and subject them to a frequency analysis. The main assumptions are that the data are independent of each other and are from a stationary time series. A stationary time series is a data series collected over time and having statistical properties that do not change over time.

An intuititve estimate for the magnitude of frequent floods can be made based on our understanding of the concept of return period. For ex[illegible]le, the 5-year flood is one that is equaled or exceeded on the ave[illegible]e once every five years or about 20 percent of the time. [illegible] peaks exceed 2120 cfs. Therefore, we might estimate the magnitude of the 5-year flood as 2120 cfs. Similarly 10 percent of the flows exceed 2400 cfs so we can estimate the 10-year event as 2400 cfs.

A difficulty with this intuitive approach is that the magnitude of events having return periods longer than the length of the available record cannot be estimated. Also the magnitude of events having return periods close to the record length is dependent on very few observations and is thus somewhat uncertain. For example, the 10-year event in the above example depends on only 3 observations. What is needed is a procedure for utilizing all of the data to describe the probabilistic nature of the peak flows.

A start in this direction can be made by plottting the data in the form of a frequency histogram. This is merely a plot of the frequency of occurrence of peak flows in some class interval versus the class interval. Figure 2.6 is such a plot using a class interval of 750 cfs. Similarly a plot of the percent of the values greater than or equal to a given value versus the magnitude of the value can be made. Figure 2.7 is a plot of this nature for the Beargrass Creek data. From Figure 2.7 the magnitude of the 5-year flood (p = 1/T = 1/5 = 0.20 or 20 percent chance of occurrence) can be estimated as about 2150 cfs and the 10-year flood (10 percent chance of occurrence) is about 3250 cfs.

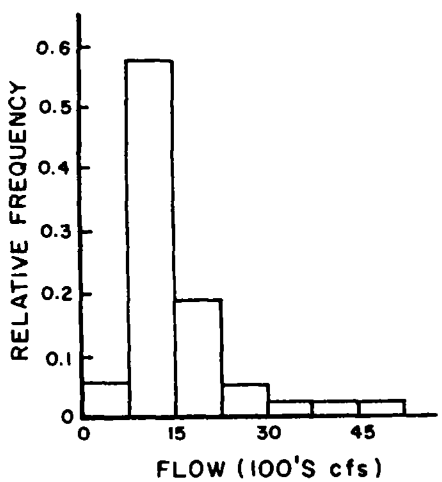

Figure 2.6. Frequency histogram – Beargrass Creek data.

When considerable data are available, this is a reasonable procedure to use for estimating low return period floods. Inspection of Figure 2.7 shows that the data exhibit some 'roughness' and that perhaps a better estimate for low return period floods could be obtained by drawing a smooth curve through the data and then using the curve to define the magnitude of floods with various return periods.

Unfortunately a plot such as Figure 2.7 is generally not sufficient for estimating the magnitude of a longer return period flood. For example, the 25-year flood can be determined from Figure 2.7 by reading the smooth curve at the 4 percent point. This is not a very reliable estimate, however, because it depends almost entirely on the magnitude of the two largest events in the record. If the largest flood event in the record had been 7000 cfs or 4200 cfs or some other value, this would have greatly altered our estimate for the 25-year flood.

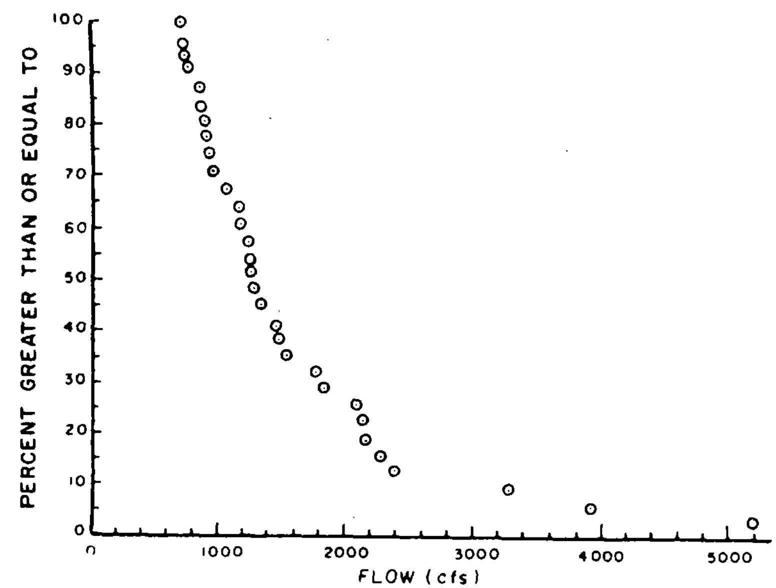

Figure 2.7. Empirical flood-frequency curve – Beargrass Creek data.

Furthermore, the estimation of a 100-year flood based on this data requires the smooth curve be extrapolated to the one percent point. This extrapolation, and indeed the entire smooth curve, would be extremely dependent on the whims of the individual doing the extrapolation. Different individuals would estimate different values for the 100-year flood and the values could differ by several thousand cfs.

What is needed is an analytic method for placing a curve through the plotted points. This analytic curve could then be used to estimate the magnitude of floods with various return periods. Before discussing analytic techniques for flood frequency analysis, the matter of plotting random data (flood peaks) requires further attention.

The procedure arrived at in preparing Figure 2.7 results in the point 707 cfs being plotted at the 100 percent point or we are stating that 100 percent of all annual flood peaks on this stream will be greater than 707 cfs. Even though this is true for the particular 31-year record that is available, we do not know that it is true for all time and would suspect that there is a chance that in some future year an annual flood peak of less than 707 cfs might occur. Thus we would like to avoid assigning a 100 percent chance or probability of 1 to any event.

A second consideration in plotting flood peaks against probability is that when arithmetic graph paper is used as in Figure 2.7, the points generally form an extremely curved pattern with the larger flood widely spaced. To overcome this inconvenience, special paper known as probability paper has been developed. Several kinds of probability paper are available. The most widely available are normal probability paper and lognormal probability paper. Lognormal probability paper will be used in this treatment.

The steps to be followed in plotting random data on probability paper are:

1. Rank the data from the largest to the smallest.
2. Calculate the plotting position from:

$$p = m/(n+1) \qquad (2.8)$$

where p is the plotting position, m is the rank of the observation and n is the number of years of data.

3. Plot the observation on probability paper with p along the probability scale and magnitude along the variable scale.

As an example of probability plotting, consider the Beargrass Creek data. This data is ranked and the plotting positions determined in Table 2.3. Figure 2.8 is a plot of the data on lognormal probability paper. Since the data were ranked from the largest to the smallest, the plotting position, p, represents the fraction of the values greater than or equal to the corresponding value of the data. The data do not plot as a straight line on lognormal paper, but the curvature is greatly reduced over that shown in Figure 2.7.

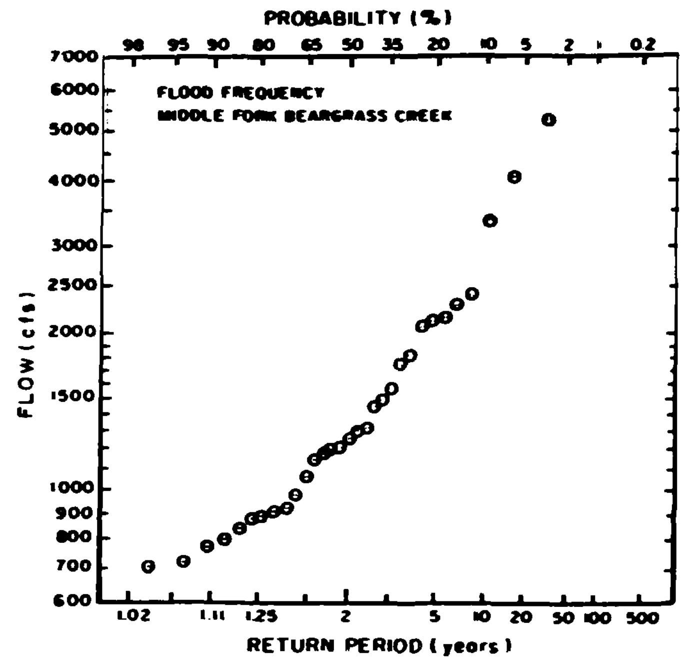

Figure 2.8. Probability plot – Beargrass Creek data.

At this point a smooth curve can be sketched through the data or we can use analytical methods to 'fit' a line through the points. In this latter approach, an equation having unknown parameters is used to describe the data much like the straight line y=a+bX is fitted through plotted points on regular graph paper. The difficulty we now face is selecting the 'equation' to use and in estimating the parameters of this equation.

Table 2.3 Plotting Position – Middle Fork Beargrass Creek, Cannons Lane, Louisville, Kentucky.

Year	Discharge	Rank	Plotting Position
1945	1810	9	0.281
1946	791	28	0.875
1947	839	27	0.844
1948	1750	10	0.313
1949	898	24	0.750
1950	2120	7	0.219
1951	1220	18	0.563
1952	1290	15	0.469
1953	768	29	0.906
1954	1570	11	0.344
1955	1240	17	0.531
1956	1060	21	0.656
1957	1490	12	0.375
1958	884	25	0.781
1959	1320	14	0.438
1960	3300	3	0.094
1961	2400	4	0.125
1962	976	22	0.688
1963	918	23	0.719
1964	3920	2	0.063
1965	1150	20	0.625
1966	874	26	0.813
1967	712	30	0.938
1968	1450	13	0.406
1969	707	31	0.969
1970	5200	1	0.031
1971	2150	6	0.188
1972	1170	19	0.594
1973	2080	8	0.250
1974	1250	16	0.500
1975	2270	5	0.156

Equations for describing the probability of occurrence of random events are known as probability density functions (pdf) and cumulative distribution functions (cdf). A pdf can be used to evaluate the probability of a random event in a specified interval. A cdf can be used to evaluate the probability of an event being equal to or less than a given value. We will use the notation $p_X(x)$ and $P_X(x)$ to ... of the random variable X evaluated at X = x.

These two are related by

$$P_X(x) = \int_{-\infty}^{x} p_X(t)\,dt \tag{2.9}$$

where X is the random variable and t is a variable of integration. There are a limitless number of functions that can be used for pdf's. The only requirements for a function to be a pdf are:

1. $p_X(x) \geq 0$ for all x
2. $\int_{-\infty}^{\infty} p_X(x)\,dx = 1$

Pdf's may take on any number of shapes. The most familiar is the bell-shaped curve of the normal probability density function shown in Figure 2.9. The normal pdf is given by

$$p_X(x) = (\sqrt{2\pi}\,\sigma_X)^{-1}\, e^{-\frac{1}{2}(x-\mu_X)^2/\sigma_X^2} \tag{2.10}$$

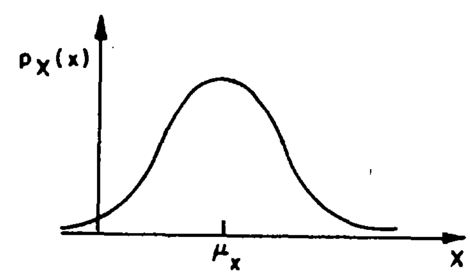

Figure 2.9. Normal distribution.

The normal distribution is symmetrical about the mean μ_X and ranges from $-\infty$ to ∞. The normal distribution is generally not used in flood frequency determinations because it permits negative values and because flood frequency distributions are generally not symmetrical. For example, the Beargrass Creek data in Figure 2.6 exhibits a pronounced tailing off to the right which is typical of flood peak data. Even though the normal distribution is generally not used in flood frequency analyses, we will continue to consider it since an understanding of it is essential for statistical work.

The cdf of the normal distribution is

$$P_X(x) = \int_{-\infty}^{x} (\sqrt{2\pi}\,\sigma_X)^{-1}\, e^{-\frac{1}{2}(t-\mu_X)^2/\sigma_X^2}\,dt \tag{2.11}$$

which gives the probability that $X \leq x$.

$$P_X(x) = \text{prob}(X \leq x) \quad (2.12)$$

The probability that X is between a and b can be evaluated from:

$$\text{prob}(a \leq X \leq b) = \text{prob}(X \leq b) - \text{prob}(X \leq a) \quad (2.13)$$

$$= P_X(b) - P_X(a)$$

$$= \int_a^b p_X(t)\, dt$$

The normal distribution is a two-parameter distribution with the parameters being μ_X, the mean X and σ_X, the standard deviation of X. For any application of the normal distribution we must estimate μ_X and σ_X by $\bar{X}$ and S_X.

Using equations 2.4 and 2.5 the mean and standard deviation of the Beargrass Creek data are found to be 1599 cfs and 1006 cfs respectively. Now *IF* the normal distribution was an adequate representation of the Beargrass Creek data, it could be used to make probabilistic statements concerning the data. For example, the probability of a peak less than or equal to 2500 cfs could be evaluated as

$$\text{prob}(Q \leq 2500) = P_Q(2500)$$

$$= \int_{-\infty}^{2500} (\sqrt{2\pi}\ 1006)^{-1}\, e^{-\frac{1}{2}(t-1599)^2/1006^2}\, dt \quad (2.14)$$

Unfortunately this latter expression cannot be analytically evaluated and numerical procedures must be used. To overcome the problem of requiring a separate numerical integration for the normal distribution for every possible combination of the parameters μ_X and σ_X, a transformation of variables is made using

$$Z = (X - \mu_X)/\sigma_X \quad (2.15)$$

Z is called a standardized random variable. The expression

$$p_Z(z) = (2\pi)^{-\frac{1}{2}}\, e^{-Z^2/2} \quad (2.16)$$

is known as the standard normal distribution. Equation 2.14 can now be evaluated as

$$\text{prob}(Q \leq x) = \text{prob}(Z \leq (x - \mu_X)/\sigma_X) \quad (2.17)$$

or

$$\text{prob}(Q \leq 2500) = \text{prob}(Z \leq (2500 - 1599)/1006)$$

$$= \text{prob}(Z \leq 0.896) = \int_{-\infty}^{0.896} (2\pi)^{-\frac{1}{2}}\, e^{-Z^2/2}\, dZ$$

The latter expression can be evaluated using tables of the standard normal distribution. Care must be exercised in using these table to see exactly what information the tables contain.

Table 2.1A of Appendix 2A contains prob $(0 \leq Z \leq z)$. Using thi table prob $(0 \leq Z \leq 0.896)$ is found to be 0.314. Since prob $(Z \leq 0.896)$ = prob $(-\infty \leq Z \leq 0.00)$ + prob $(0 \leq Z \leq 0.896)$ and prob $(-\infty \leq Z \leq 0.00)$ is 0.500, the desired probability is 0.814.

The interpretation of this calculation is that if the flood peak on Beargrass Creek can be described by a normal distribution with a mean of 1599 cfs and a standard deviation of 1006 cfs, then 81.[illegible] percent of the annual peaks should be less than or equal to 2500 cfs. The data tabulation actually shows that 28 of the 31 values of 90.[illegible] percent are less than or equal to 2500 cfs.

Looking back at equation 2.13, it is apparent that prob$(a \leq X \leq b)$ is the area under the pdf, $p_X(x)$, between X = a and X = b. Thus, the probability of a random observation falling in the interval a to b is the area under the pdf between a and b. In a sense the relative frequency histogram of Figure 2.6 gives similar information. Based on the data in hand, we would estimate for example, that the probability that a random annual peak would fall in the interval 1500 cfs to 2250 cfs is 0.190. There is apparently a relationship between relative frequency and probability. Denote by $f_X(x_i)$ the relative frequency of observations in an interval of width Δx centered on x_i. The probability of an observation falling in this interval is

$$\text{prob}(x_i - \Delta x/2 \leq X \leq x_i + \Delta x/2) = \int_{x_i - \Delta x/2}^{x_i + \Delta x/2} p_X(x)\, dx \quad (2.18)$$

which is the area under $p_X(x)$ between $x_i - \Delta x/2$ and $x_i + \Delta x/2$. This area can be approximated by $\Delta x\ p_X(x_i)$ which is the width of the interval times the height of $p_X(x)$ evaluated at x_i (Figure 2.10).

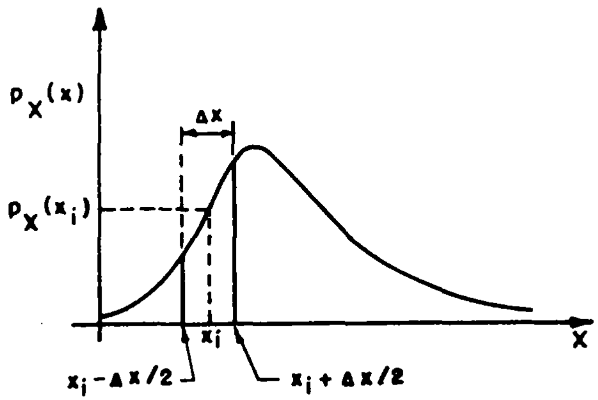

Figure 2.10. Calculation of prob($X_i - \Delta x/2 < X < X_i + \Delta x/2$).

Therefore, the relationship between the relative frequency of observations in an interval Δx and the pdf is

$$f_X(x_i) = \Delta x\ p_X(x_i) \qquad (2.19)$$

Since probability is related to the area under the pdf, it is apparent that prob(X = x) for a continuous random variable must be zero since

$$\text{prob}(X = x) = \int_x^x p_X(t)\,dt = 0 \qquad (2.20)$$

We can use equation 2.19 and Figure 2.6 to visually judge the appropriateness of using the normal distribution to describe the Beargrass Creek data. Table 2.4 shows, under the assumption of a normal distribution, the observed and expected frequency of observations in several classes. The data are plotted in Figure 2.11. Entries in the expected relative frequency column of Table 2.4 are based on Equation 2.19 and the normal distribution. For example, for the second class

$$f_X(1125) = 750(\sqrt{2\pi}\ 1006)^{-1}\ e^{-\frac{1}{2}(1125 - 1599)^2/1006^2} = 0.267$$

Table 2.4 Observed and Expected Frequency – Beargrass Creek Data (normal distribution).

Class interval	Observed rel. freq.	Expected rel. freq.
0- 750	0.064	0.141
750-1500	0.581	0.267
1500-2250	0.194	0.286
2250-3000	0.064	0.177
3000-3750	0.032	0.063
3750-4500	0.032	0.012
4500-5250	0.032	0.001
	0.999	0.947

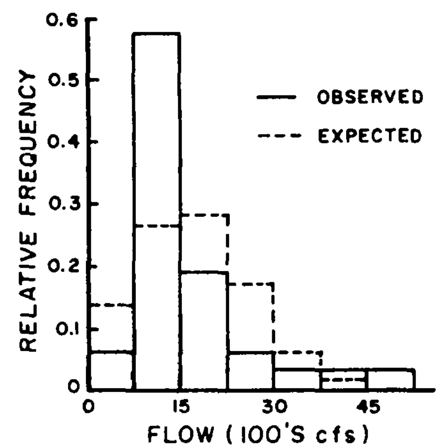

Figure 2.11. Comparison of observed and expected flow frequency (under assumptio of observations in class intervals) – Beargrass Creek data.

A second visual comparison between the observed data and th assumed distribution (the normal distribution) can be made by usi the normal distribution as the equation for the line describing the da in Figure 2.7. Equation 2.11 can be used to draw a line through t points of Figure 2.7 assuming the points are from a normal distrib tion. All that is required is to calculate the prob($Q \geq q$) for the vario values of Q, and then plot this probability versus q. The prob($Q >$ is equal to 1 - prob($Q \leq q$) since prob($Q = q$) is zero and Q mu either be $<q$ or $>q$. To get prob($Q > q$) we first evaluate prob($Q \leq$ Equations 2.14 and 2.17 show such a calculation for Q = 2500 c

Table 2.5 shows the results of similar calculations for several values of Q. The prob(Q > q) is plotted in Figure 2.12.

Table 2.5 Comparison of Observed and Expected Cumulative Probabilities (normal distribution).

Q	Observed % ≥ Q	Expected % ≤ Q	Expected % ≥ Q
700	100.0	18.7	81.3
1000	67.7	28.8	71.2
1500	35.5	46.0	54.0
2000	25.8	65.6	34.4
2500	9.6	81.6	18.4
3000	9.6	91.8	8.2
4000	3.2	99.2	0.8
5000	3.2	100.0	0.0
6000	0	100.0	0.0

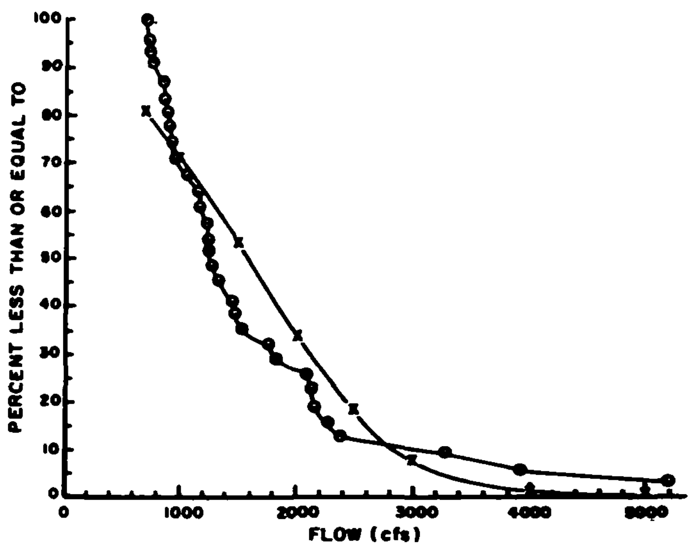

Figure 2.12. Comparison of flood frequency – Beargrass Creek data – observed and normal distributions.

From either Figure 2.11 or 2.12 it is apparent that the normal distribution is not a satisfactory approximation to the observed data of Beargrass Creek. Another probability distribution must be found to describe the data. This involves finding another m[illegible]ematical function to use as a pdf and cdf in place of equations 2.1[illegible] and 2.11

used for the normal distribution. There are available a large number of such expressions. Again, these expressions are known as probability distributions.

The three probability distributions that receive the most attention for describing flood frequencies are the lognormal (LN), extreme value type I (EVI) and log Pearson type III (LP3). This treatment will be restricted to these three distributions. Other distributions are discussed in Haan (1977).

As noted earlier, the Beargrass Creek data, when plotted in the form of a relative frequency histogram, tailed off to the right much more than it did to the left. This tailing off to the right results in a positive skewness. The normal distribution is symmetrical about the mean, and as such, has a skewness of zero. The LN, EVI and LP3 distributions can all accommodate positively skewed data.

For the LN distribution, the skewness γ and the coefficient of variation C_v are related by

$$\gamma = 3C_v + C_v^3 \tag{2.21}$$

For the EVI distribution, γ is a constant 1.139. There are no restrictions on γ for the LP3 distribution.

Expressions can be written for the pdf's and cdf's for the LN, EVI and LP3 distributions. They can then be considered in a manner analogous to the treatment given to the normal distribution. This is not necessary, however, since simpler methods are available. Reference can be made to Haan (1977) for a detailed treatment of these and other distributions.

Chow (1951) has shown that many types of frequency analyses can be reduced to

$$X_T = \bar{X}(1 + C_v K_T) \tag{2.22}$$

where X_T is the magnitude of the event with return period T, $\bar{X}$ is the mean of the original data, C_v is the coefficient of variation of the original [illegible]a and K_T is a frequency factor which is a function of the probability distribution selected and properties of the original data.

The frequency factors for the LN distribution as a function of C_v are contained in Table 2.6. Table 2.7 contains the frequency factors for the EVI distribution. All that is required for selecting K_T for this distribution is knowledge of the sample size and the desired return period.

Table 2.6 Frequency Factors for Lognormal Distribution. (Chow, 1964)

Return Period					Corresponding C_v
1.01	2	5	20	100	
-2.33	0	0.84	1.64	2.33	0
-2.25	-0.02	0.84	1.67	2.40	0.033
-2.18	-0.04	0.83	1.70	2.47	0.067
-2.11	-0.06	0.82	1.72	2.55	0.100
-2.04	-0.07	0.81	1.75	2.62	0.136
-1.98	-0.09	0.80	1.77	2.70	0.166
-1.91	-0.10	0.79	1.79	2.77	0.197
-1.85	-0.11	0.78	1.81	2.84	0.230
-1.79	-0.13	0.77	1.82	2.90	0.262
-1.74	-0.14	0.76	1.84	2.97	0.292
-1.68	-0.15	0.75	1.85	3.03	0.324
-1.63	-0.16	0.73	1.86	3.09	0.351
-1.58	-0.17	0.72	1.87	3.15	0.381
-1.54	-0.18	0.71	1.88	3.21	0.409
-1.49	-0.19	0.69	1.88	3.26	0.436
-1.45	-0.20	0.68	1.89	3.31	0.462
-1.41	-0.21	0.67	1.89	3.36	0.490
-1.38	-0.22	0.65	1.89	3.40	0.517
-1.34	-0.22	0.64	1.89	3.44	0.544
-1.31	-0.23	0.63	1.89	3.48	0.570
-1.28	-0.24	0.61	1.89	3.52	0.596
-1.25	-0.24	0.60	1.89	3.55	0.620
-1.22	-0.25	0.59	1.89	3.59	0.643
-1.20	-0.25	0.58	1.88	3.62	0.667
-1.17	-0.26	0.57	1.88	3.65	0.691
-1.15	-0.26	0.56	1.88	3.67	0.713
-1.12	-0.26	0.55	1.87	3.70	0.734
-1.10	-0.27	0.54	1.87	3.72	0.755
-1.08	-0.27	0.53	1.86	3.74	0.776
-1.06	-0.27	0.52	1.86	3.76	0.796
-1.04	-0.28	0.51	1.85	3.78	0.818
-1.01	-0.28	0.49	1.84	3.81	0.857
-0.98	-0.29	0.47	1.83	3.84	0.895
-0.95	-0.29	0.46	1.81	3.87	0.930
-0.92	-0.29	0.44	1.80	3.89	0.966
-0.90	-0.29	0.42	1.78	3.91	1.000
-0.84	-0.30	9.39	1.75	3.93	1.081
-0.80	-0.30	0.37	1.71	3.95	1.155

Table 2.7 Frequency Factors for Extreme Value Type I Distribution.

Sample size n	5	10	15	20	25	50	75	100	1000
15	0.967	1.703	2.117	2.410	2.632	3.321	3.721	4.005	6.265
20	0.919	1.625	2.023	2.302	2.517	3.179	3.563	3.836	6.006
25	0.888	1.575	1.963	2.235	2.444	3.088	3.463	3.729	5.842
30	0.866	1.541	1.922	2.188	2.393	3.026	3.393	3.653	5.727
35	0.851	1.516	1.891	2.152	2.354	2.979	3.341	3.598	
40	0.838	1.495	1.866	2.126	2.326	2.943	3.301	3.554	5.576
50	0.820	1.466	1.831	2.086	2.283	2.889	3.241	3.491	5.478
60	0.807	1.446	1.806	2.059	2.253	2.852	3.200	3.446	
70	0.797	1.430	1.788	2.038	2.230	2.824	3.169	3.413	5.359
80	0.788	1.417	1.773	2.020	2.212	2.802	3.145	3.387	
90	0.782	1.409	1.762	2.007	2.198	2.785	3.125	3.367	
100	0.779	1.401	1.752	1.998	2.187	2.770	3.109	3.349	5.261
∞	0.719	1.305	1.635	1.866	2.044	2.592	2.911	3.137	4.936

The steps in using the LP3 distribution are:

1. Transform the n original observation, X_i, to their logarithmic values, Y_i, by the relation

$$Y_i = \log X_i \tag{2.23}$$

2. Compute the mean logarithm, $\overline{Y}$.
3. Compute the standard deviation of the logarithms, s_y.
4. Compute the coefficient of skewness C_s from

$$C_s = \frac{n^2 \Sigma Y_i^3 - 3n \Sigma Y_i \Sigma Y_i^2 + 2(\Sigma Y_i)^3}{n(n-1)(n-2) s_y^3} \tag{2.24}$$

or

$$C_s = \frac{n \Sigma (Y_i - \overline{Y})^3}{(n-1)(n-2) s_y^3}$$

5. Compute

$$Y_T = \overline{Y} + s_Y K_T \tag{2.25}$$

where K_T is from Table 2.8. This relationship is identical to equation 2.22 except it is based on logarithms.

6. Calculate

$$X_T = \text{antilog } Y_T \tag{2.26}$$

Table 2.8 Frequency Factors For Pearson Type III Distribution.[1]

Skew Coef. C_s	Recurrence Interval in Years: 1.0101	2	5	10	25	50	100	200
3.0	-0.667	-0.396	0.420	1.180	2.278	3.152	4.051	4.970
2.8	-0.714	-0.384	0.460	1.210	2.275	3.114	3.973	4.847
2.6	-0.769	-0.368	0.499	1.238	2.267	3.071	3.889	4.718
2.4	-0.832	-0.351	0.537	1.262	2.256	3.023	3.800	4.584
2.2	-0.905	-0.330	0.574	1.284	2.240	2.970	3.705	4.444
2.0	-0.990	-0.307	0.609	1.302	2.219	2.912	3.605	4.298
1.8	-1.087	-0.282	0.643	1.318	2.193	2.848	3.499	4.147
1.6	-1.197	-0.254	0.675	1.329	2.163	2.780	3.388	3.990
1.4	-1.318	-0.225	0.705	1.337	2.128	2.706	3.271	3.828
1.2	-1.449	-0.195	0.732	1.340	2.087	2.626	3.149	3.661
1.0	-1.588	-0.164	0.758	1.340	2.043	2.542	3.022	3.489
.8	-1.733	-0.132	0.780	1.336	1.993	2.453	2.891	3.312
.6	-1.880	-0.099	0.800	1.328	1.939	2.359	2.755	3.132
.4	-2.029	-0.066	0.816	1.317	1.880	2.261	2.615	2.949
.2	-2.178	-0.033	0.830	1.301	1.818	2.159	2.472	2.763
0	-2.326	0	0.842	1.282	1.751	2.054	2.326	2.576
-.2	-2.472	0.033	0.850	1.258	1.680	1.945	2.178	2.388
-.4	-2.615	0.066	0.855	1.231	1.606	1.834	2.029	2.201
-.6	-2.755	0.099	0.857	1.200	1.528	1.720	1.880	2.016
-.8	-2.891	0.132	0.856	1.166	1.448	1.606	1.733	1.837
-1.0	-3.022	0.164	0.852	1.128	1.366	1.492	1.588	1.664
-1.2	-3.149	0.195	0.844	1.086	1.282	1.379	1.449	1.501
-1.4	-3.271	0.225	0.832	1.041	1.198	1.270	1.318	1.351
-1.6	-3.388	0.254	0.817	0.994	1.116	1.166	1.197	1.216
-1.8	-3.499	0.282	0.799	0.945	1.035	1.069	1.087	1.097
-2.0	-3.605	0.307	0.777	0.895	0.959	0.980	0.990	0.995
-2.2	-3.705	0.330	0.752	0.844	0.888	0.900	0.905	0.907
-2.4	-3.800	0.351	0.725	0.795	0.823	0.830	0.832	0.833
-2.6	-3.889	0.368	0.696	0.747	0.764	0.768	0.769	0.769
-2.8	-3.973	0.384	0.666	0.702	0.712	0.714	0.714	0.714
-3.0	-4.051	0.396	0.636	0.660	0.666	0.666	0.667	0.667

1. Water Resources Council (1967).

EXAMPLE ANALYTICAL FREQUENCY ANALYSIS

As an example of applying these three distributions, again consider the data of Table 2.2. The mean and standard deviation of the original data were found to be 1599 and 1006 respectively. The C_v is 0.629. Values of K_T for various values of T for the log-normal distribution are selected from Table 2.6, and equation 2.22 gives the corresponding values of X_T. These results are shown in Table 2.9.

Table 2.9 Flood Frequency Analysis for Middle Fork, Beargrass Creek, Cannons Lane, Louisville, Kentucky.

	Return Period (years) 5	25	100
lognormal distribution			
K_T	.60	2.11	3.57
X_T	2202	3718	5190
extreme value type I			
K_T	.863	2.385	3.642
X_T	2467	3998	5284
log Pearson type III			
K_T	.772	2.011	2.937
Y_T	7.628	8.257	8.726
X_T	2056	3853	6161

Values of K_T for various return periods for the extreme value type I distribution are selected from Table 2.7 and X_T again comes from equation 2.22. These results are shown in Table 2.9.

In applying the log Pearson type III distribution, $\overline{Y}$ was found to be 7.237, s_y is 0.507 and C_s from equation 2.24 is 0.87. K_T values are then selected from Table 2.8, Y_T calculated from equation 2.25 and X_T from equation 2.26. The results of these calculations are shown in Table 2.9.

This example shows that the distribution which is selected can have a substantial affect on the estimated flood magnitude for a given frequency. This is especially apparent for the longer return periods. The best fitting lines according to the three distributions used are shown in Figure 2.13.

From Figure 2.13 it is apparent that it is difficult to select which of the three distributions best describe this data. Considering the

EVI is 1.139 while it is 2.13 for this data. For the LN Υ and C_v are related through equation 2.21. The estimated value of Υ is 2.14 which agrees quite well with the requirement of the LN. One discouraging factor concerning the LN is that Υ of the logarithms should be zero for the LN distribution while for this data it is 0.87. Apparently the LN is not a precise approximation for the data.

The skewness cannot be used to make decisions concerning the LP3 since the LP3 is a 3-parameter distribution which uses Υ to estimate these parameters. Looking at Figure 2.13, it does appear that the LP3 is a better approximation to the data at the upper end of the frequency curve.

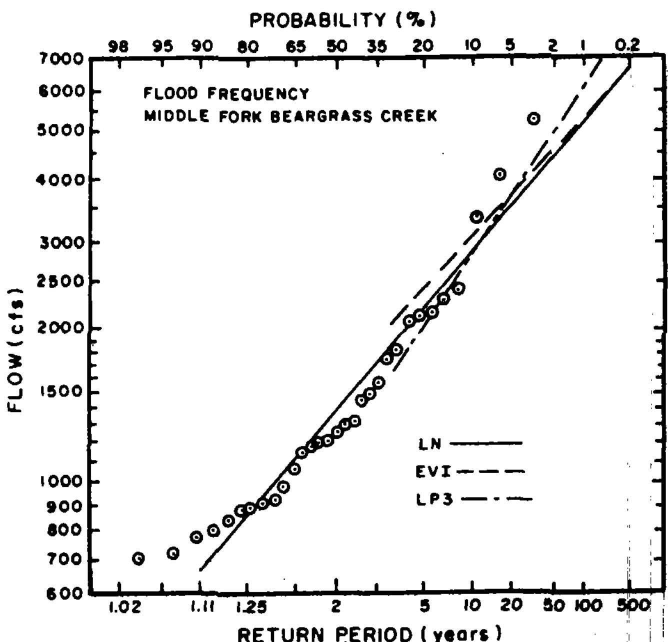

Figure 2.13. Comparison of several probability distributions for the Beargrass Creek data.

As discussed later, one problem with this data is the possible nonstationarity of peak flows due to changing watershed conditions. The difficulties experienced with this set of data demonstrated why many have been led to the recommendation that single short records are not reliable, and regional frequency analysis procedures as dis-

In summary, the Case I situation is where a relatively long record of peak flows is available on the stream of interest at the point of interest. The method of analysis is to select a probability density function to describe the data, extract certain statistics from the data, and use equation 2.22 along with appropriate frequency factors to estimate flows within the desired return period.

Some of the assumptions made in the Case I situation are:

1. The data are sufficient in quantity and quality to produce reliable estimates for the parameters of the probability distribution selected.
2. The flow characteristics of the stream have not been changing over time (stationary data series).
3. The peak flow observations are statistically independent from year to year.
4. The data are representative of the flow behavior expected during the life of the project being considered.

Assumption 4 merely extends the stationarity assumption to future flows. In watersheds with changing land use, assumptions 2 and 4 are especially troublesome in that the changing land use alters streamflow characteristics. Many times this nonhomogeneity in streamflow data is very subtle and only becomes apparent over a long period of time or when sudden and large scale changes occur. For example, Figure 2.14 is a plot of the 31-year record for the Beargrass Creek data. The bulk of the peak flows are in the range 750 to 2000 cfs. It appears as though the frequency of occurrence of peaks in excess of 2000 cfs is increasing with time. The random nature of the data makes it difficult to make firm statements in this regard.

In closing the discussion on the Case I flood frequency analysis, a word of caution is offered concerning the extrapolation of frequency data to estimate the magnitude of an event with a return period much greater than the period of record. In looking at Figure 2.13, it appears that the extrapolation of the frequency lines from the 31-year record to estimate the 100-year or even 500-year event is not much of an extrapolation. In the sense of the physical distance on the probability paper, the extrapolation is not very great; however, in the sense of extrapolating the data to 3 or possibly 15 times its original length, it is a very significant extrapolation. The nature of random data makes an extrapolation of this kind very dangerous and produces estimates of low reliability or which possess considerable uncertainty. Haan (1977) gives a procedure for evaluating the uncertainty that is present.

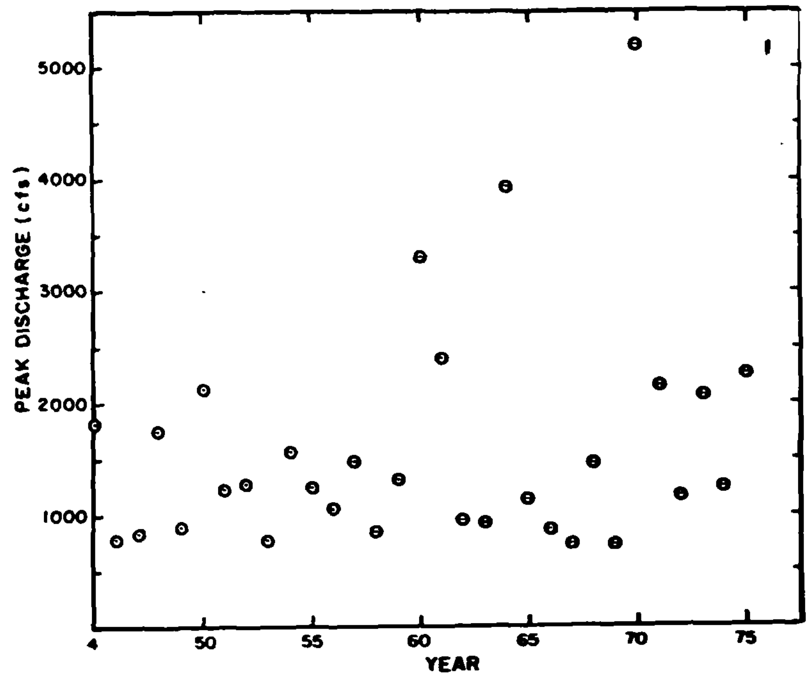

Figure 2.14. Changing flow condition resulting in nonhomogeneous time series.

To illustrate this point, a simulation was made assuming a lognormal distribution with a mean of 1599 cfs and a standard deviation of 1006 cfs. The procedure was to randomly select 15 observations from this lognormal distribution and then use these 15 observations to estimate the magnitude of the 100-year event. The 100-year event was estimated based on the lognormal distribution using the mean and standard deviation of the 15 generated values. This entire process was repeated 100 times producing 100 estimates for the 100-year flood. The probabilistic behavior of these 100 estimates was then analyzed.

The following tabulation shows the frequency of occurrence of the estimates for the 100-year flood. The true value is 5190 cfs.

The mean estimated value was 4945 cfs. If we take the 100-year flood as 5000 cfs, the following tabulation can be used to get some idea of the possible errors involved in using a short record to estimate a rare flood.

Flow (cfs)	No. of Estimates
2500-3000	7
3000-3500	8
3500-4000	13
4000-4500	10
4500-5000	21
5000-5500	15
5500-6000	7
6000-6500	5
6500-7000	6
7000-7500	3
7500-8000	2
8000-8500	2
14000- ∞	1
Total	100

Note that only 36 values or about 1/3 of the estimates are within 10 percent (500 cfs) of the actual 100-year flood. Nearly 1/2 (47 values) of the estimates are in error by more than 20 percent. 15 percent of the values are in error by more than 40 percent and 1 estimate was actually *high* by a factor of nearly 3. These numbers illustrate what can happen when a short record (15 years in this case) is used to estimate a rare flood (100-year flood in this case). This is the reason that extrapolation is dangerous. In any real situation the variability would be even greater than shown here, because the true underlying probability distribution would not be known. We assumed a lognormal distribution in this example and based our calculations on this assumption as though it were the population distribution.

CASE II - FLOOD FREQUENCY DETERMINATIONS

In some instances flow data are available on the stream of interest but at a location some distance either upstream or downstream from the point of interest. In this situation there are several procedures that can be used to estimate the flood frequency relationship at the point of interest. One method is to perform a flood frequency analysis on the available record as described in Case I, and then adjust the record to the point of interest.

The adjustment of the flow record from one location on a stream to another can be done in a number of ways. If the flow record includes the entire flood hydrograph, this hydrograph, could be routed to the point of interest making the proper adjustments for local inflows and outflows along the routing reach. This method is a very good one, but requires more data than is generally available and is quite time consuming.

A second method of flow adjustment is to correlate flood peaks with drainage basin characteristics, and then use this correlation to adjust the flow rate. The most common characteristic used in a situation like this is the basin area. Quite frequently the peak discharge for a given frequency is related to the basin area by an equation of the form

$$Q_T = a\,A^b \qquad (2.27)$$

where Q_T is the T-year flood magnitude, A is the basin area, and a and b are constants. The coefficient b generally ranges from 0.5 to 1.0.

If data on a stream are available at 2 locations, the coefficients can be estimated from that data. For example, consider the data in Table 2.10. Here information on the 2-year, 5-year and 10-year floods is available at two locations. The exponents a and b of equation 2.27 can be estimated for each return period by substitution. Consider the 10-year data:

$$Q_{10} = a\,A^b$$

(i) $3300 = a(2.3)^b$ for the larger basin
(ii) $1100 = a(0.7)^b$ for the smaller basin

The ratio of equations i and ii

$$3300/1100 = (2.3/0.7)^b$$

can be solved for b using logarithms resulting in b = 0.923. Substituting this estimate for b into equation i results in

$$3300 = a\,(2.3)^{0.923}$$

which gives an estimate for a of 1527. Similar calculations for the 2- and 5-year floods result in the estimates shown in Table 2.10. From this an average value of b is 0.92. The coefficient a is seen to be a function of return period T.

Table 2.10 Hypothetical Flood Frequency Data.

T	Area (mi^2) 0.7	2.3		
years	Q_T (cfs)		a	b
2	750	2250	1041	0.923
5	950	2800	1318	0.909
10	1100	3300	1527	0.923

To estimate the 10-year flood at a point on the stream where the drainage area is 1.5 square miles, equation 2.27 and the estimated 10-year coefficients are used.

$$Q_{10} = 1527(1.5)^{0.92} = 2217 \text{ cfs}$$

If data at only one other location on the stream is available, one can still estimate the flow at a second location *if* an assumption as to the coefficient b is made. For small differences in area, a reasonable assumption for b is 1.0. For example, considering the 10-year data for the 0.7 mi^2 basin only and taking b as 1.0 results in an estimate for a of

$$a = 1100/0.7 = 1571$$

The 10-year flood can then be estimated on the 1.5 mi^2 basin as

$$Q_{10} = 1571(1.5)^1 = 2357 \text{ cfs}$$

When these approaches

exercised. First and most importantly, the basic flood producing characteristics of all the basins must be the same. There cannot be a mix of drastically differing land uses unless some type of land use variable is included in the prediction equation for a and b. Returning to the above example, it has been assumed that all three watersheds are similar. On surface mined watersheds this is a severe limitation and generally means that there cannot be a very large difference in the areas of watersheds considered or the method of mining employed.

If the watershed characteristics are changing along the stream, then calculations such as shown here can be used as an aid in estimating a flow of a given return period, but will not give the final estimate. If the available data represent a mixture of mined and undisturbed conditions and our point of interest is below only the mined part of the basin, the flow estimated by a straight forward application of the techniques presented here will most likely have to be adjusted to reflect the fact that it is below a mined area, while the coefficients a and b were estimated based on a partly undisturbed basin.

CASE III - FLOOD FREQUENCY DETERMINATION

It is not uncommon to find that a streamflow record at the point of interest may be too short to use in a flood frequency analysis. This may be the result of a gaging program that was only recently changed so that much of the earlier portion of the streamflow record is no longer representative of the basin.

A short streamflow record can be a great aid in checking calculations and procedures used in flood frequency estimation. If a major drainage project is to be planned, the local governing body would be wise to install a stream gage early in the feasibility part of the project planning process. In this way, by the time the final design is made, some streamflow data would be available. For relatively large drainage projects, this short-term gaging approach is relatively inexpensive and can easily pay for itself through the re-

A short streamflow record is one of less than 10 years in length. A record such as this will contain a great deal of information, but will be insufficient for a Case I frequency analysis. Presumably, a record of the rainfall that produced the recorded runoff will be available or can be estimated from nearby raingages. These records on rainfall and runoff can now be used to estimate the empirical coefficients in an approximate model. The model might be a continuous simulation model, an event or hydrograph model, or a model for estimating peak flow only. The type of model selected will depend on the quantity and quality of available data and the purpose of the analysis.

Regardless of the type of model selected, the model will have empirical coefficients associated with it that must be estimated. In general, the number of empirical coefficients required is proportional to the complexity of the model with the continuous simulation models having the most coefficients. The importance of actual data on the stream of interest in estimating the empirical coefficients of these models cannot be overemphasized.

Once the model coefficients have been estimated, a long-term rainfall record can be processed through the model to produce a long-term streamflow record. This simulated, long-term streamflow record can then be subjected to a frequency analysis as discussed under Case I if necessary.

In the event that a long-term rainfall record is not available for the site under study, one can use records from the nearest raingage. Fortunately, in many parts of the country, rainfall records can be transferred a few miles without introducing significant affects on the estimated peak runoff rates from major events. The long-term records from nearby raingages may not be usable for runoff parameter estimation since the recorded rainfall may have been considerably different from what actually fell on the watershed. The record can be used for simulation, however, because the long-term statistical properties may be the same as those of rains that actually fell on the watershed.

In the absence of any applicable long-term rainfall records, it duce a synthetic rainfall record to use in simulations with the run-off model.

Another approach to using the information contained in a short record is to use the short record to estimate a low return period index flood. This might be the 2-year flood for example. Knowing the magnitude of this index flood, a regionalized relationship between the ratio of the index flood and floods of a greater return period can be used to estimate the magnitude of less frequent events. The determination of the regionalized relationship is covered under the Case IV situation discussed below.

CASE IV - FLOOD FREQUENCY DETERMINATION

Often one finds that streamflow records are available at several nearby locations while none or a very short record is available at the point of interest. Several methods for using the information on the nearby stations to augment whatever is known at the site of interest are available. These methods generally fall under the heading of 'Regional Flood Frequency Analysis'.

One widely used method of regional flood frequency analysis is discussed by Dalrymple (1960). The method consists of computing a base flood frequency relationship in terms of the return period and the ratio of the peak flow for a given return period to an index flood (usually the mean annual flood) for several streams in the region. The median value of this ratio is then plotted versus the return period. Figure 2.15 is such a plot for 18 stations in Alberta and Saskatchewan, Canada, as reported by Durant and Blackwell (1959).

The first step in this regional approach is to select several streamflow records from nearby locations that are 'hydrologically similar' to the basin of interest. At each of these locations, a Case I flood frequency analysis is made. An index flood is then defined. This might be the 2-year flood. The ratio of the magnitude of the T-year flood to the index flood is computed for several values of T at each location. The ratio is then plotted versus T and a smooth curve drawn through the points (Figure 2.15).

The next step is to relate the index flood to watershed characteristics. The area of the watershed is generally used along with other

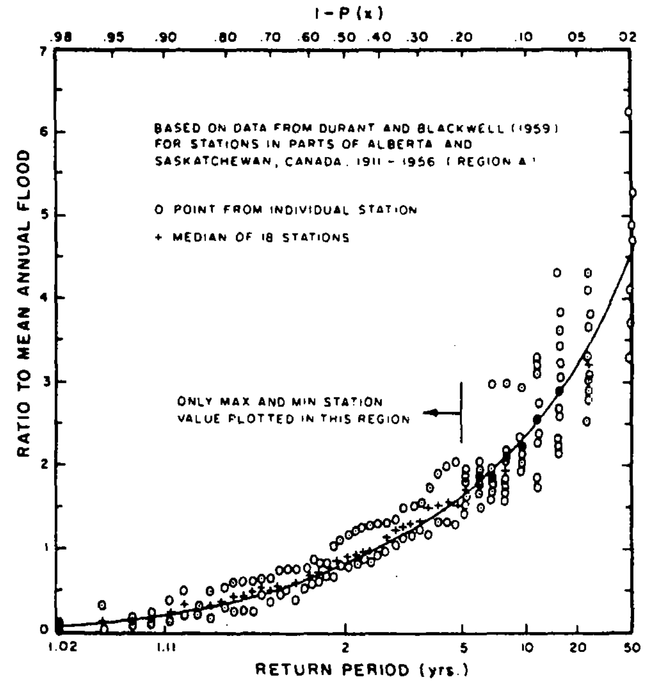

Figure 2.15a. Regional flood frequency analysis.

geomorphic, physical and meteorological factors. This step is generally done through a regression analysis to produce an equation of the form

$$Q_I = aX_1^{\,b} X_2^{\,c} \ldots X_n^{\,q} \qquad (2.28a)$$

or

$$Q_I = a + bX_1 + cX_2 + \ldots + qX_n \qquad (2.28b)$$

where Q_I is the magnitude of the index flood; a, b, ..., q are estimated coefficients; and X_1, X_2, ..., X_n are the watershed and climatic factors. Regressions of this type are discussed in more detail by Haan (1977).

The third step is to use equation 2.28 to estimate the index flood for the location of interest. Alternatively, if a short record is available at the location of interest, the index flood can be estimated from that record. The final step is to use the regional flood frequency curve and the estimated Q_I to calculate Q_T for the desired values of T.

A variation of the above technique for regional flood frequency analysis is to estimate Q_T for several values of T at each gaged location as explained above, and then relate Q_T to watershed factors and climatic data by regression to produce an equation like 2.28 with Q_I replaced by Q_T. A separate equation is needed for each value of T. These equations can then be used to estimate the desired value of Q_T at the study location. One disadvantage of this approach is the possibility of not retaining the proper relation between the Q_T's for different values of T. That is, one could conceivably estimate Q_{25} as being less than Q_{10}. Haan (1977) discussed the use of multivariate multiple regression to overcome this difficulty.

CASE V - FLOOD FREQUENCY DETERMINATION

When there are no streamflow records available on the stream of interest or on nearby streams, one is forced to use some type of empirical procedure for estimating the magnitude of runoff events of the desired frequency. We are now out of the realm of frequency analysis in the sense of determining the frequency of occurrence of events based on a probabilistic analysis of data.

For this situation a hydrologic model of some type must be employed. The model may range in complexity from the Rational Equation to a complete, continuous simulation model. The type of model selected will depend on the data available for model fitting, the familiarity of the person selecting the model with various models, the purpose of the modeling effort, the time and money available for completing the modeling effort, and the importance of flow estimates.

If a continuous simulation model is selected and several years of streamflow are generated, this generated data can be subjected to a flood frequency analysis. If an event-based model (such as a unit

each year can be analyzed, with these data subsequently subjected to a flood frequency analysis.

If an approach like the Rational Equation is used, one is assuming that the frequency of the estimated flow peak is the same as the frequency of the rainfall used in the equation. This is not a bad assumption over the long run. For individual events, the return period of the rainfall and the resulting runoff are not necessarily the same because of the effect of such factors as antecedent soil water content. However, over the long run, the expected or average return period of the runoff will nearly equal the return period of the rainfall.

Since the Case V situation is really a modeling effort or requires the use of hydrologic techniques not generally thought of as being frequency analysis, its treatment will be deferred to later sections of this text.

DISCUSSION OF FLOOD FREQUENCY DETERMINATIONS

The backbone of any frequency analysis is the procedure described under Case I where a particular probability distribution is selected for describing a set of data. The parameters of this distribution are then estimated and the magnitude of events for various return period computed. Methods for plotting the observed data on probability paper and for drawing in the best fitting line according to the selected distribution have also been discussed.

In the frequency analysis of streamflow data, care must be taken to see that the data are representative of the flow situation that is expected to exist during the life of the project being considered. This is a real problem for watersheds having large and hydrologically significant land use changes.

There is no set guideline on how much data is required in order for a valid frequency analysis to be made. At least 15 years of data and limited extrapolation of the data are desirable. Extrapolation is estimating the magnitude of floods having return periods considerably longer than the available period of record. Many times one is forced to do this. One must be aware of the uncertainty that exists in the resulting flow estimates.