CHAPTER 2

FREQUENCY ANALYSIS

2-1. Definition.

a. Frequency. Many of the statistical techniques that are applied to hydrologic data (to enable inferences to be made about particular attributes of the data) can be labeled with the term "frequency analysis" techniques. The term "frequency" usually connotes a count (number) of events of a certain magnitude. To have a perspective of the importance of the count, the total number of events (sample size) must also be known. Sometimes the number of events within a specified time is used to give meaning to the count, e.g., two daily flows were this low in 43 years. The probability of a certain magnitude event recurring again in the future, if the variable describing the events is continuous, (as are most hydrologic variables), is near zero. Therefore, it is necessary to establish class intervals (arbitrary subdivisions of the range) and define the frequency as the number of events that occur within a class interval. A pictorial display of the frequencies within each class interval is called a histogram (also known as a frequency polygon).

b. Relative Frequency. Another means of representing the frequency is to compute the relative frequency. The relative frequency is simply the number of events in the class interval divided by the total number of events:

$$f_i = n_i/N \qquad (2\text{-}1)$$

where:

f_i = relative frequency of events in class interval i

n_i = number of events in interval i

N = total number of events

A graph of the relative frequency values is called a frequency distribution or histogram, Figure 2-1a. As the number of observations approaches infinity and the class interval size approaches zero, the enveloping line of the frequency distribution will approach a smooth curve. This curve is termed the probability density function (Figure 2-1a).

c. Cumulative Frequency. In hydrologic studies, the probability of some magnitude being exceeded (or not exceeded) is usually the primary interest. Presentation of the data in this form is accomplished by accumulating the probability (area) under the probability density function. This curve is termed the cumulative distribution function. In most statistical texts, the area is accumulated from the smallest event to the largest. The accumulated area then represents non-exceedance probability or percentage (Figure 2-1b). It is more common in hydrologic studies to accumulate the area from the largest event to the smallest. Area accumulated in this manner represents exceedance probability or percentage.

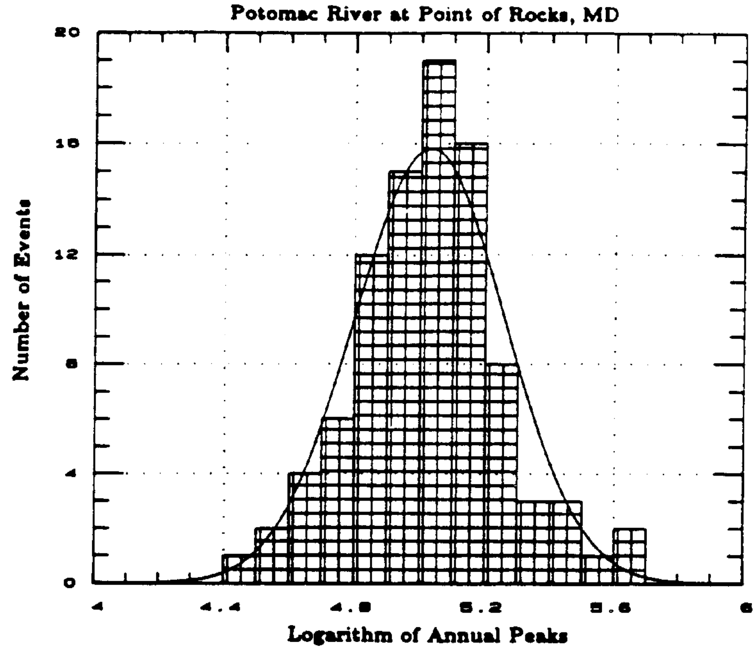

Figure 2-1a. Histogram and Probability Density Function.

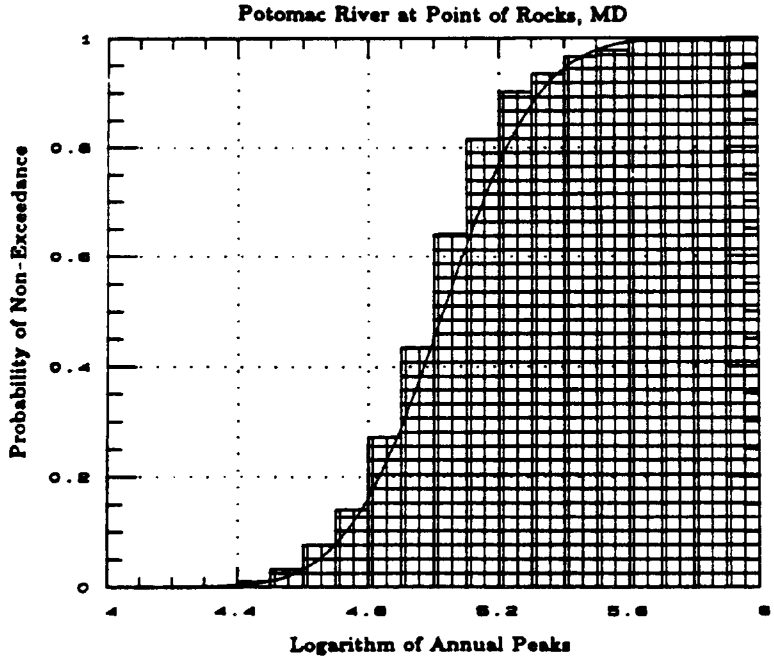

Figure 2-1b. Sample and Theoretical Cumulative Distribution Functions.

2-2. Duration Curves.

a. Computation. The computation of flow-duration curves was probably the first attempt to analyze hydrologic data by statistical techniques. The events for flow-duration curves are usually mean daily flow values. One of the first steps in preparation of a duration curve is dividing the range of the data into class intervals. Table 2-1 shows the class intervals of daily flows input into the computer program STATS (58) for a duration analysis of Fishkill Creek at Beacon, New York. The flows tend to be grouped near the low end with very few large flows. Therefore, the relative frequency curve is skewed to the right. It has been found that making the logarithmic transform reduces the skewness of the curve. The class intervals in Table 2-1 are based on a logarithmic distribution of the flows. Plotting the data in Table 2-1 on log-probability paper, Figure 2-2, provides a plot that is easily read at the extremities of the data. The daily flow-duration curve cannot be considered a frequency curve in the true sense, because the daily flow on a particular day is highly correlated with the flow on the preceding day. For this reason, the abscissa is labeled "percent of time."

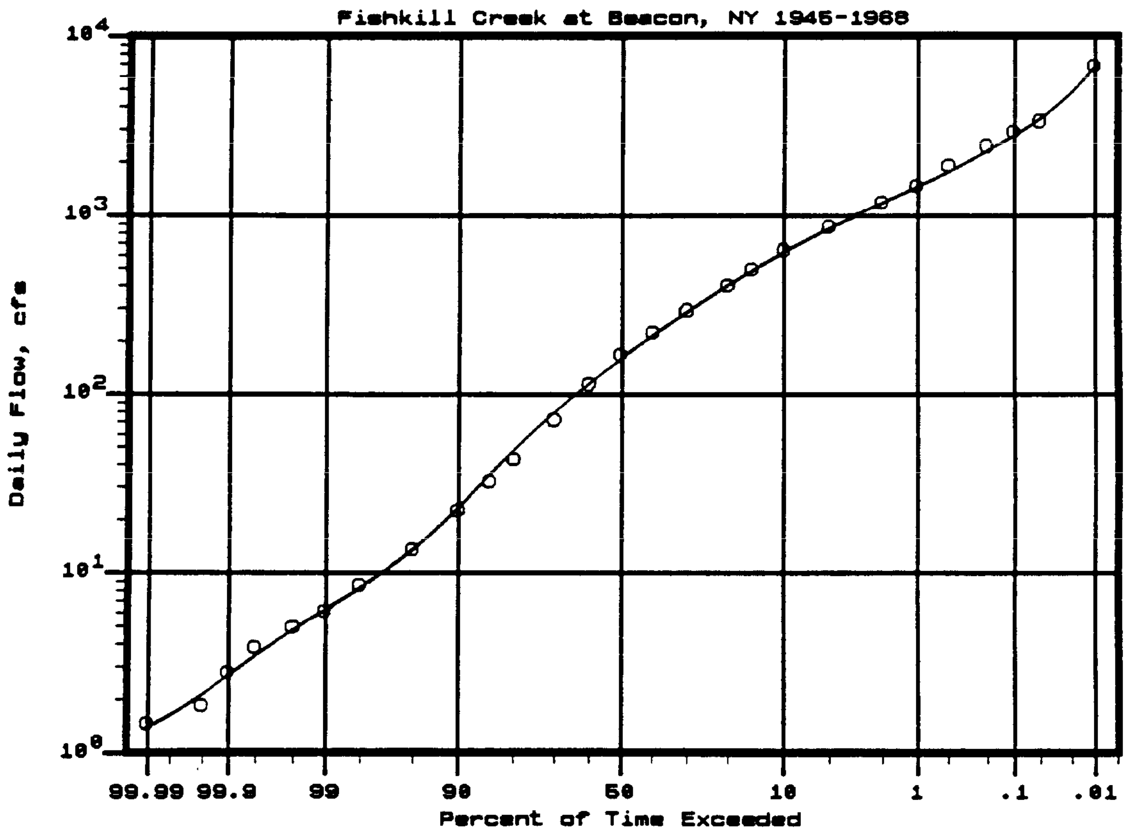

Figure 2-2. Daily Flow-Duration Curve.

Table 2-1. Daily Flow-Duration Data and Interpolated Values.

-DURATION DATA- FISHKILL CR AT BEACON, NY - DAILY FLOWS

CLASS NUMBER	LOWER CLASS LIMIT FLOW,CFS	NUMBER IN CLASS	ACCUM NUMBER	PERCENT EQUAL OR EXCEED	CLASS NUMBER	LOWER CLASS LIMIT FLOW,CFS	NUMBER IN CLASS	ACCUM NUMBER	PERCENT EQUAL OR EXCEED
1	1.00	5	8766	100.00	16	100.00	888	5606	63.95
2	2.00	4	8761	99.94	17	150.00	783	4718	53.82
3	3.00	8	8757	99.90	18	200.00	1246	3935	44.89
4	4.00	22	8749	99.81	19	300.00	822	2689	30.68
5	5.00	37	8727	99.56	20	400.00	484	1867	21.30
6	6.00	66	8690	99.13	21	500.00	340	1383	15.78
7	8.00	95	8624	98.38	22	600.00	465	1043	11.90
8	10.00	254	8529	97.30	23	800.00	239	578	6.59
9	15.00	261	8275	94.40	24	1000.00	251	339	3.87
10	20.00	423	8014	91.42	25	1500.00	47	88	1.00
11	30.00	405	7591	86.60	26	2000.00	32	41	0.47
12	40.00	359	7186	81.98	27	3000.00	6	9	0.10
13	50.00	332	6827	77.88	28	4000.00	0	3	0.03
14	60.00	480	6495	74.09	29	6000.00	3	3	0.03
15	80.00	409	6015	68.62	30	7000.00	0	0	0.00

-INTERPOLATED DURATION CURVE- FISHKILL CR AT BEACON, NY - DAILY FLOWS

PERCENT EQUAL OR EXCEED	INTERPOLATED MAGNITUDE FLOW,CFS	PERCENT EQUAL OR EXCEED	INTERPOLATED MAGNITUDE FLOW,CFS
0.01	6970.0	60.00	118.0
0.05	3480.0	70.00	74.5
0.10	3020.0	80.00	44.7
0.20	2530.0	85.00	33.4
0.50	1960.0	90.00	22.7
1.00	1500.0	95.00	14.0
2.00	1230.0	98.00	8.8
5.00	903.0	99.00	6.3
10.00	658.0	99.50	5.2
15.00	518.0	99.80	4.0
20.00	420.0	99.90	2.9
30.00	306.0	99.95	1.9
40.00	230.0	99.99	1.5
50.00	171.0	100.00	1.1

Output from HEC computer program STATS.

b. Uses. Duration curves are useful in assessing the general low flow characteristics of a stream. If the lower end drops rapidly to the probability scale, the stream has a low ground-water storage and, therefore, a low or no sustained flow. The overall slope of the flow-duration curve is an indication of the flow variability in the stream. Specific uses that have been made of duration curves are: 1) assessing the hydropower potential of run-of-river plants; 2) determining minimum flow release; 3) water quality studies; 4) sediment yield studies; and 5) comparing yield potential of basins. It must be remembered that the chronology of the flows is lost in the assembly of data for duration curves. For some studies, the low-flow sequence, or persistence, may be more important (see Chapter 4).

c. Monthly Curves. Occasionally the distribution of the flows during particular seasons of the year is of interest. Figure 2-3 illustrates a way of presenting daily-flow-duration curves that were computed from the daily flows during each month.

d. Stage-Duration. Stage-duration curves are often used to establish vertical navigation clearances for bridges. If there have been no changes in the discharge versus stage relationship (rating curve), then the stages may be used instead of flows to compute a stage-duration curve. But, if there have been significant changes to the rating curve (because of major levee construction, for instance) then the stage-duration curve should be derived from the flow-duration curve and the latest rating curve. The log transformation is not recommended for stages.

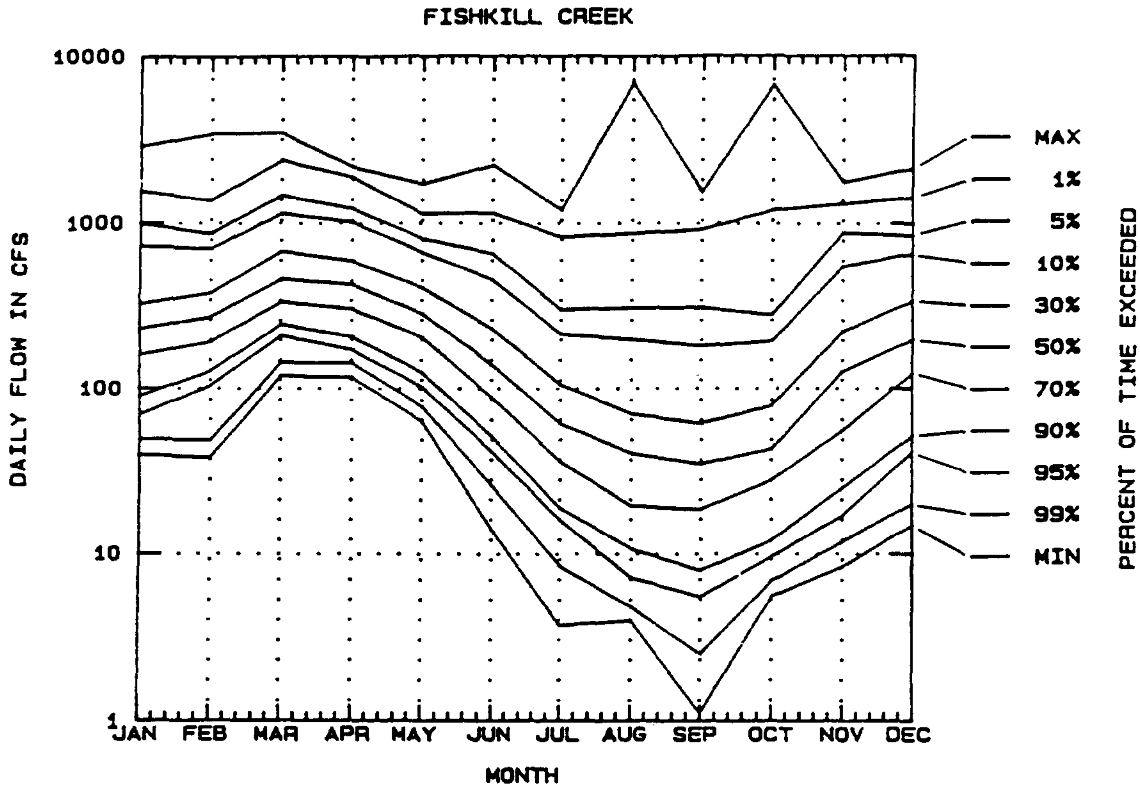

Figure 2-3. Daily Flow-Duration Curves for Each Month.

e. Future Probabilities. A duration curve is usually based on a fairly large sample size. For instance, Figure 2-2 is based on 8766 daily values (Table 2-1). Even though the observed data can be used to make inferences about future probabilities, conclusions drawn from information at the extremities can be misleading. The data indicate there is a zero percent chance of exceeding 6970 cfs, however, it is known that there is a finite probability of experiencing a larger flow. And similarly, there is some chance of experiencing a lower flow than the recorded 1.1 cfs. Therefore, some other means is needed for computing the probabilities of infrequent future events. Section 2-4 describes the procedure for assigning probabilities to independent events.

2-3. Selection of Data for Frequency Analysis.

a. Selection Based on Application of Results. The primary question to be asked before selection of data for a frequency study is: how will the frequency estimates be used? If the frequency curve is to be used for estimating damage that is related to the peak flow in a stream, maximum peak flows should be selected from the record. If the damage is best related to a longer duration of flow, the mean flow for several days' duration may be appropriate. For example, a reservoir's behavior may be related to the 3-day or 10-day rain flood volume or to the seasonal snowmelt volume. Occasionally, it is necessary to select a related variable in lieu of the one desired. For example, where mean-daily flow records are more complete than the records of peak flows, it may be more desirable to derive a frequency curve of mean-daily flows and then, from the computed curve, derive a peak-flow curve by means of an empirical relation between mean daily flows and peak flows. All reasonably independent values should be selected, but only the annual maximum events should be selected when the application of analytical procedures discussed in Chapter 3 is contemplated.

b. Uniformity of Data.

(1) General Considerations. Data selected for a frequency study must measure the same aspect of each event (such as peak flow, mean-daily flow, or flood volume for a specified duration), and each event must result from a uniform set of hydrologic and operational factors. For example, it would be improper to combine items from old records that are reported as peak flows but are in fact only daily readings, with newer records where the instantaneous peak was actually measured. Similarly, care should be exercised when there has been significant change in upstream storage regulation during the period of record to avoid combining unlike events into a single series. In such a case, the entire record should be adjusted to a uniform condition (see Sections 2-3f and 3-9). Data should always be screened for errors. Errors have been noted in published reports of annual flood peaks. And, errors have been found in the computer files of annual flood peaks. The transfer of data to either paper or a computer file always increases the probability that errors have been accidentally introduced.

(2) Mixed Populations. Hydrologic factors and relationships during a general winter rain flood are usually quite different from those during a spring snowmelt flood or during a local summer cloudburst flood. Where two or more types of floods are distinct and do not occur predominately in mutual combinations, they should not be combined into a single series for frequency analysis. It is usually more reliable in such cases to segregate the data in accordance with type and to combine only the final curves, if necessary. In the Sierra Nevada region of California and Nevada, frequency studies are made separately for rain floods which occur principally during the months of November through March,

and for snowmelt floods, which occur during the months of April through July. Flows for each of these two seasons are segregated strictly by cause - those predominantly caused by snowmelt and those predominantly caused by rain. In desert regions, summer thunderstorms should be excluded from frequency studies of winter rain flood or spring snowmelt floods and should be considered separately. Along the Atlantic and Gulf Coasts, it is often desirable to segregate hurricane floods from nonhurricane events. Chapter 10 describes how to combine the separate frequency curves into one relation.

c. Location Differences. Where data recorded at two different locations are to be combined for construction of a single frequency curve, the data should be adjusted as necessary to a single location, usually the location of the longer record. The differences in drainage area, precipitation and, where appropriate, channel characteristics between the two locations must be taken into account. When the stream-gage location is different from the project location, the frequency curve can be constructed for the stream-gage location and subsequently adjusted to the project location.

d. Estimating Missing Events. Occasionally a runoff record may be interrupted by a period of one or more years. If the interruption is caused by destruction of the gaging station by a large flood, failure to fill in the record for that flood would result in a biased data set and should be avoided. However, if the cause of the interruption is known to be independent of flow magnitude, the record should be treated as a broken record as discussed in Section 3-2b. In cases where no runoff records are available on the stream concerned, it is usually best to estimate the frequency curve as a whole using regional generalizations, discussed in Chapter 9, instead of attempting to estimate a complete series of individual events. Where a longer or more complete record at a nearby station exists, it can be used to extend the effective length of record at a location by adjusting frequency statistics (Section 3-7) or estimating missing events through correlation (Chapter 12).

e. Climatic Cycles. Some hydrologic records suggest regular cyclic variations in precipitation and runoff potential, and many attempts have been made to demonstrate that precipitation or streamflows evidence variations that are in phase with various cycles, particularly the well established 11-year sunspot cycle. There is no doubt that long-duration cycles or irregular climatic changes are associated with general changes of land masses and seas and with local changes in lakes and swamps. Also, large areas that have been known to be fertile in the past are now arid deserts, and large temperate regions have been covered with glaciers one or more times. Although the existence of climatic changes is not questioned, their effect is ordinarily neglected, because the long-term climatic changes have generally insignificant effect during the period concerned in water development projections, and short-term climatic changes tend to be self-compensating. For these reasons, and because of the difficulty in differentiating between stochastic (random) and systematic changes, the effect of natural cycles or trends during the analysis period is usually neglected in hydrologic frequency studies.

f. Effect of Basin Development on Frequency Relations.

(1) Hydrologic frequency estimates are often used for some purpose relating to planning, design or operation of water resources control measures (structural and nonstructural). The anticipated effects of these measures in changing the rate and volume of flow is assessed by comparing the without project frequency curve with the corresponding with project frequency curve. Also, projects that have existed in the past have affected the rates and volumes of flows, and the recorded values must be adjusted to reflect uniform conditions in order that the frequency analysis will conform to the basic

assumption of homogeneity. In order to meet the assumptions associated with analytical frequency analysis techniques, the flows must be essentially unregulated by manmade storage or diversion structures. Consequently, wherever practicable, recorded runoff values should be adjusted to natural (unimpaired) conditions before an analytical frequency analysis is made. In cases where the impairment results from a multitude of relatively small improvements that have not changed appreciably during the period of record, it is possible that analytical frequency analysis techniques can be applied. The adjustment to natural conditions may be unnecessary and, because of the amount of work involved, not cost effective.

(2) One approach to determining a frequency curve of regulated or modified runoff consists of routing all of the observed flood events under conditions of proposed or anticipated development. Then a relationship is developed between the modified and the natural flows, deriving an average or dependable relationship. A frequency curve of modified flows is derived from this relationship and the frequency curve of natural flows. In order to determine frequencies of runoff for extreme floods, routings of multiples of the largest floods of record or multiples of a large hypothetical flood can be used. Techniques of estimating project effects are outlined in Chapter 3-09d.

g. Annual Series Versus Partial Duration Series. There are two basic types of frequency curves used to estimate flood damage. A curve of annual maximum events is ordinarily used when the primary interest lies in the larger events or when the second largest event in any year is of little concern in the analysis. The partial-duration curve represents the frequency of all independent events of interest, regardless of whether two or more occurred in the same year. This type of curve is sometimes used in economic analysis, where there is considerable damage associated with the second largest and third largest floods that occurred in some of the years. Caution must be exercised in selecting events because they must be both hydrologically and economically independent. The selected series type should be established early in the study in coordination with the planner and/or economist. The time interval between flood events must be sufficient for recovery from the earlier flood. Otherwise damage from the later flood would not be as large as computed. When both the frequency curve of annual floods and the partial-duration curve are used, care must be exercised to assure that the two are consistent. A graphic demonstration of the relation between a chronologic record, an annual-event curve and a partial-duration curve is shown on Figure 2-4.

h. Presentation of Data and Results of Frequency Analysis. When frequency curves are presented for technical review, adequate information should be included to permit an independent review of the data, assumptions and analysis procedures. The text should indicate clearly the scope of the studies and include a brief description of the procedure used, including appropriate references. A summary of the basic data consisting of a chronological tabulation of values used and indicating sources of data and any adjustments should be included. The frequency data should also be presented in graphical form, ordinarily on probability paper, along with the adopted frequency curves. Confidence limit curves should also be included for analytically-derived frequency curves to illustrate the relative value of the frequency relationships. A map of the gage locations and tables of the adopted statistics should also be included.

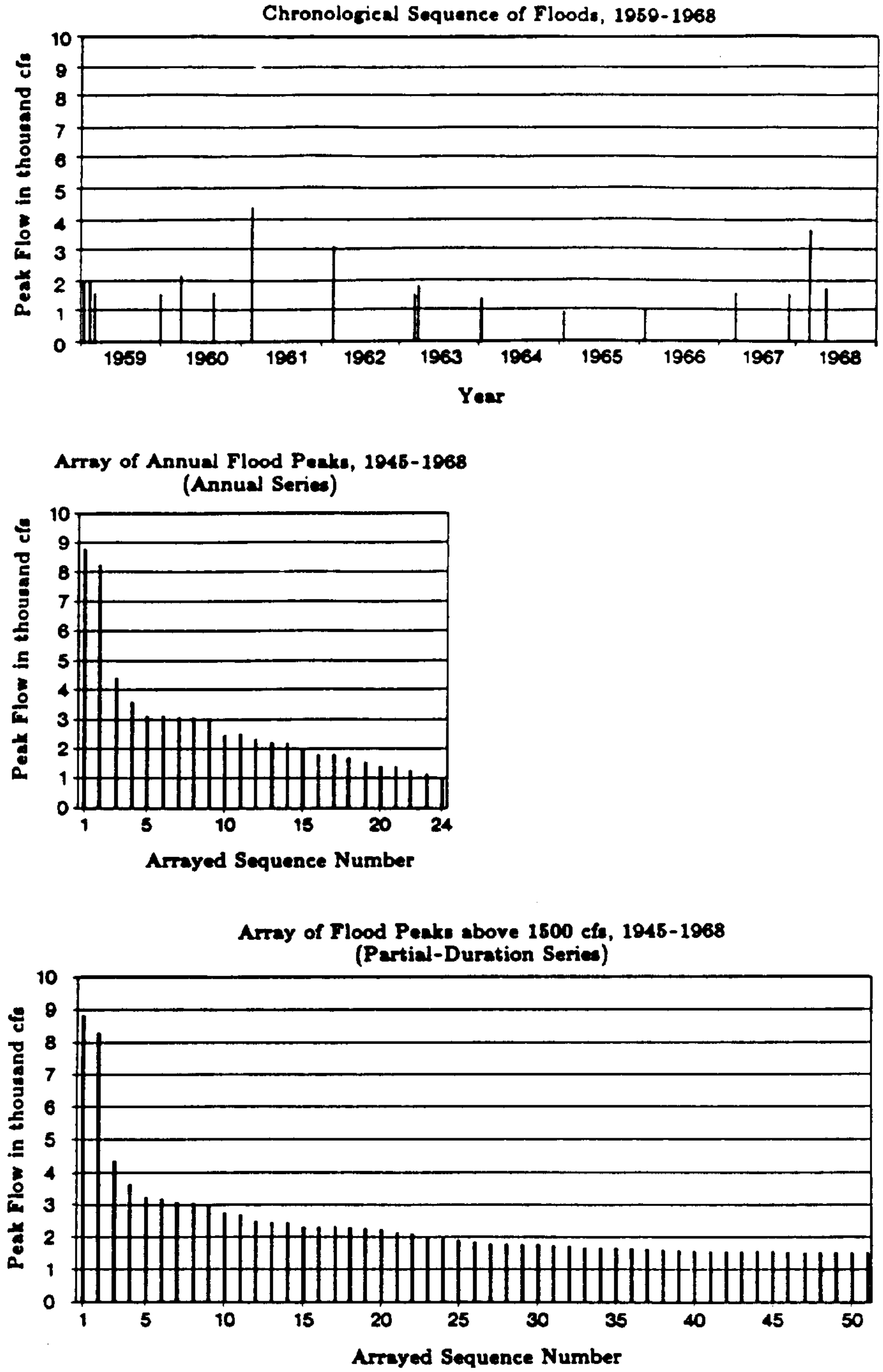

Figure 2-4. Illustration of Chronologic Sequence and Arrayed Flood Peaks.

2-4. Graphical Frequency Analysis.

a. Advantages and Limitations. Every set of frequency data should be plotted graphically, even though the frequency curves are obtained analytically. It is important to visually compare the observed data with the derived curve. The graphical method of frequency-curve determination can be used for any type of frequency study, but analytical methods have certain advantages when they are applicable. The principal advantages of graphical methods are that they are generally applicable, that the derived curve can be easily visualized, and that the observed data can be readily compared with the computed results. However, graphical methods of frequency analysis are generally less consistent than analytical methods as different individuals would draw different curves. Also, graphical procedures do not provide means for evaluating the reliability of the estimates. Comparison of the adopted curve with plotted points is not an index of reliability, but it is often erroneously assumed to be, thus implying a much greater reliability than is actually attained. For these reasons, graphical methods should be limited to those data types where analytical methods are known not to be generally applicable. That is, where frequency curves are too irregular to compute analytically, for example, stream or reservoir stages and regulated flows. Graphical procedures should always be to visually check the analytical computations.

b. Selection and Arrangement of Peak Flow Data. General principles in the selection of frequency data are discussed in Section 2-3. Data used in the construction of frequency curves of peak flow consist of the maximum instantaneous flow for each year of record (for annual-event curve) or all of the independent events that exceed a selected base value (for partial-duration curve). This base value must be smaller than any flood flow that is of importance in the analysis, and should also be low enough so that the total number of floods in excess of the base equals or exceeds the number of years of record. Table 2-2 is a tabulation of the annual peak flow data with dates of occurrence, the data arrayed in the order of magnitude, and the corresponding plotting positions.

c. Plotting Formulas. Median plotting positions are tabulated in Table F-1. In ordinary hydrologic frequency work, N is taken as the number of years of record rather than the number of events, so that percent chance exceedance can be thought of as the number of events per hundred years. For arrays larger than 100, the plotting position, P_1, of the largest event is obtained by use of the following equation:

$$P_1 = 100\,(1 - (.5)^{1/N}) \qquad (2\text{-}2a)$$

The plotting position for the smallest event (P_N) is the complement ($1-P_1$) of this value, and all the other plotting positions are interpolated linearly between these two. The median plotting positions can be approximated by

$$P_m = 100(m - .3)/(N + .4) \qquad (2\text{-}2b)$$

where m is the order number of the event.

For partial-duration curves, particularly where there are more events than years (N), plotting positions that indicate more than one event per year can also be obtained using

Table 2-2. Annual Peaks, Sequential and Arrayed with Plotting Positions.

-PLOTTING POSITIONS-FISHKILL CREEK AT BEACON, N.Y.

EVENTS ANALYZED				ORDERED EVENTS			
MON	DAY	YEAR	FLOW,CFS	RANK	WATER YEAR	FLOW,CFS	MEDIAN PLOT POS
3	5	1945	2290.	1	1955	8800.	2.87
12	27	1945	1470.	2	1956	8280.	6.97
3	15	1947	2220.	3	1961	4340.	11.07
3	18	1948	2970.	4	1968	3630.	15.16
1	1	1949	3020.	5	1953	3220.	19.26
3	9	1950	1210.	6	1952	3170.	23.36
4	1	1951	2490.	7	1962	3060.	27.46
3	12	1952	3170.	8	1949	3020.	31.56
1	25	1953	3220.	9	1948	2970.	35.66
9	13	1954	1760.	10	1958	2500.	39.75
8	20	1955	8800.	11	1951	2490.	43.85
10	16	1955	8280.	12	1945	2290.	47.95
4	10	1957	1310.	13	1947	2220.	52.05
12	21	1957	2500.	14	1960	2140.	56.15
2	11	1959	1960.	15	1959	1960.	60.25
4	6	1960	2140.	16	1963	1780.	64.34
2	26	1961	4340.	17	1954	1760.	68.44
3	13	1962	3060.	18	1967	1580.	72.54
3	28	1963	1780.	19	1946	1470.	76.64
1	26	1964	1380.	20	1964	1380.	80.74
2	9	1965	980.	21	1957	1310.	84.84
2	15	1966	1040.	22	1950	1210.	88.93
3	30	1967	1580.	23	1966	1040.	93.03
3	19	1968	3630.	24	1965	980.	97.13

Output from HEC computer program HECWRC.

Equation 2-2b. This is simply an approximate method used in the absence of knowledge of the total number of events in the complete set of which the partial-duration data constitute a subset.

d. Plotting Grids. The plotting grid recommended for annual flood flow events is the logarithmic normal grid developed by Allen Hazen (ref 13) and designed such that a logarithmic normal frequency distribution will be represented by a straight line, Figure 2-5. The plotting grid used for stage frequencies is often the arithmetic normal grid. The plotting grids may contain a horizontal scale of exceedance probability, exceedance frequency, or percent chance exceedance. Percent chance exceedance (or nonexceedance) is the recommended terminology.

e. Example Plotting of Annual-Event Frequency Curve. Figure 2-5 shows the plotting of a frequency curve of the annual peak flows tabulated in Table 2-2. A smooth curve should be drawn through the plotted points. Unless computed by analytical frequency procedures, the frequency curve should be drawn as close to a straight line as possible on the chosen probability graph paper. The data plotted on Figure 2-5 shows a tendency to curve upward, therefore, a slightly curved line was drawn as a best fit line.

f. Example Plotting of Partial-Duration Curve. The partial-duration curve corresponding to the partial-duration data in Table 2-3 is shown of Figure 2-6a. This curve has been drawn through the plotted points, except that it was made to conform with the annual-event curve in the upper portion of the curve. The annual-event curve was developed in accordance with the procedures described in Chapter 3. When partial-duration data must include more events than there are years of record (see Subparagraph b) it will be necessary to use logarithmic paper for plotting purposes, as on Figure 2-6a, in order to plot exceedance frequencies greater than 100 percent. Otherwise, the curve can be plotted on probability grid, as illustrated on Figure 2-6b.

2-5. Analytical Frequency Analysis.

a. General Procedures and Common Distributions. The fitting of data by an analytical procedure consists of selecting a theoretical frequency distribution, estimating the parameters of the distribution from the data by some fitting technique, and then evaluating the distribution function at various points of interest. Some theoretical distributions that have been used in hydrologic frequency analysis are the normal (Gaussian), log normal, exponential, two-parameter gamma, three-parameter gamma, Pearson type III, log-Pearson type III, extreme value (Gumbel) and log Gumbel. Chapter 3 describes the fitting of the log-Pearson type III to annual flood peaks and Appendix C describes fitting the extreme value (Gumbel) distribution.

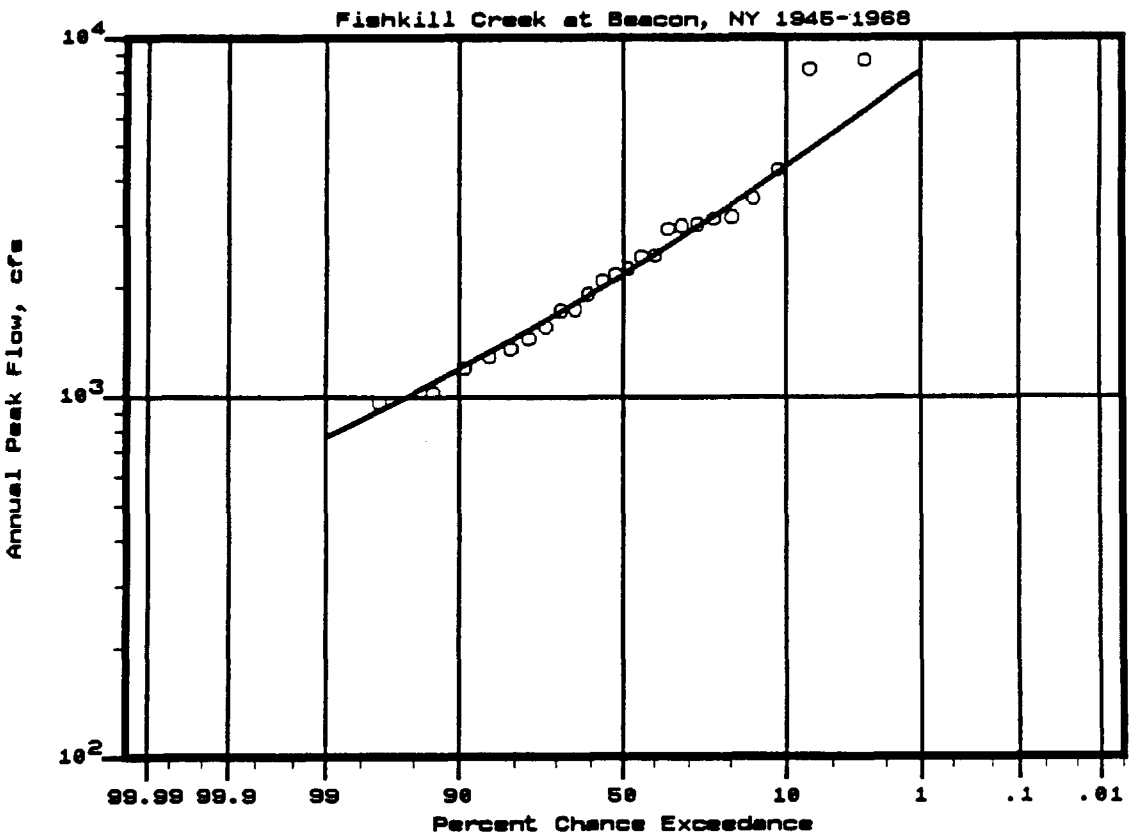

Figure 2-5. Example of Graphical Frequency Analysis

Table 2-3. Partial Duration Peaks, Arrayed with Plotting Positions.

FISHKILL CREEK AT BEACON, NY -- PEAKS ABOVE 1500 CFS

ORDERED EVENTS

RANK	WATER YEAR	FLOW,CFS	MEDIAN PLOT POS
1	1955	8800.	2.87
2	1956	8280.	6.97
3	1961	4340.	11.07
4	1968	3630.	15.16
5	1953	3220.	19.26
6	1952	3170.	23.36
7	1962	3060.	27.46
8	1949	3020.	31.56
9	1948	2970.	35.66
10	1948	2750.	39.75
11	1949	2700.	43.85
12	1958	2500.	47.95
13	1951	2490.	52.05
14	1952	2460.	56.15
15	1945	2290.	60.25
16	1953	2290.	64.34
17	1958	2290.	68.44
18	1953	2280.	72.54
19	1948	2220.	76.64
20	1951	2210.	80.74
21	1960	2140.	84.84
22	1953	2080.	88.93
23	1959	1960.	93.03
24	1959	1920.	97.13
25	1958	1900.	101.23
26	1952	1820.	105.33
27	1945	1780.	109.43
28	1963	1780.	113.52
29	1956	1770.	117.62
30	1954	1760.	121.72
31	1952	1730.	125.82
32	1968	1720.	129.92
33	1955	1660.	134.02
34	1958	1650.	138.11
35	1958	1650.	142.21
36	1953	1630.	146.31
37	1960	1610.	150.41
38	1956	1600.	154.51
39	1958	1590.	158.61
40	1958	1580.	162.70
41	1967	1580.	166.80
42	1951	1560.	170.90
43	1959	1560.	175.00
44	1955	1550.	179.10
45	1951	1540.	183.20
46	1968	1530.	187.30
47	1960	1520.	191.39
48	1958	1520.	195.49
49	1952	1520.	199.59
50	1948	1510.	203.69
51	1963	1510.	207.79

b. <u>Advantages</u>. Determining the frequency distribution of data by the use of analytical techniques has several advantages. The use of an established procedure for fitting a selected distribution would result in consistent frequency estimates from the same data set by different persons. Error distributions have been developed for some of the theoretical distributions that enable computing the degree of reliability of the frequency estimates (see Chapter 8). Another advantage is that it is possible to regionalize the parameter estimates which allows making frequency estimates at ungaged locations (see Chapter 9).

c. <u>Disadvantages</u>. The theoretical fitting of some data can result in very poor frequency estimates. For example, stage-frequency curves of annual maximum stages are shaped by the channel and valley characteristics, backwater conditions, etc. Another example is the flow-frequency curve below a reservoir. The shape of this frequency curve would depend not only on the inflow but the capacities, operation criteria, etc. Therefore, graphical techniques must be used where analytical techniques provide poor frequency estimates.

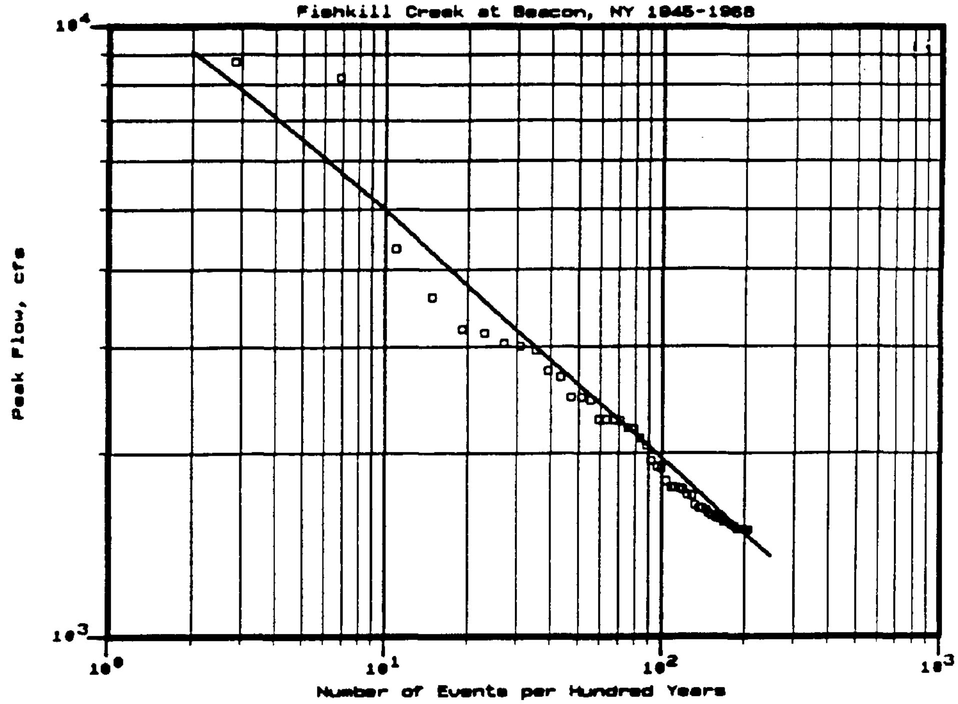

Figure 2-6a. Partial Duration Frequency Curve, Log-Log Paper

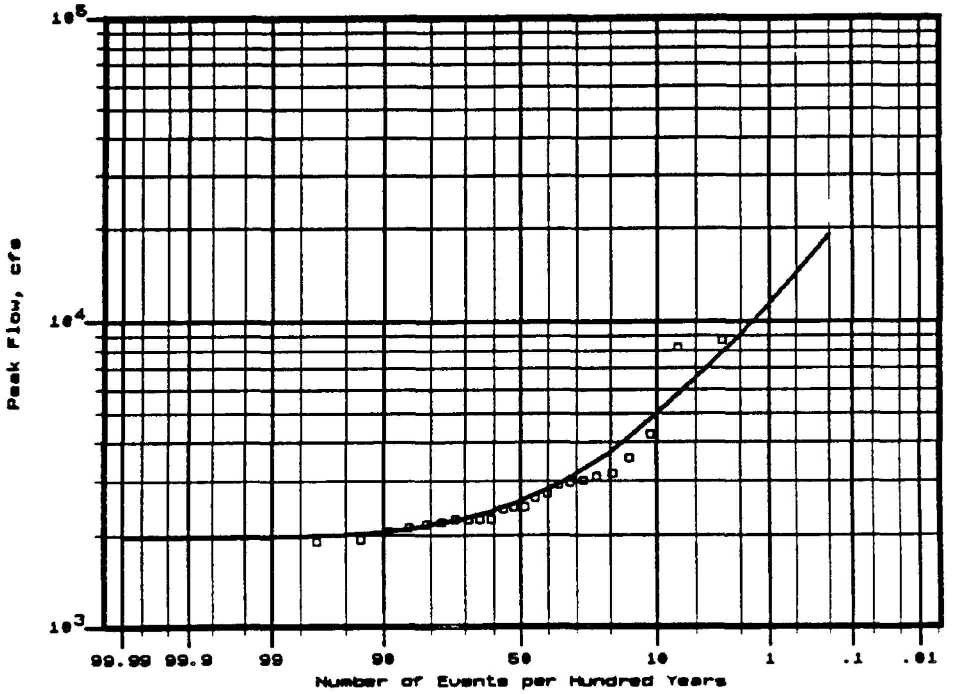

Figure 2-6b. Partial Duration Frequency Curve, Probability Paper.