US Army Corps of Engineers
Hydrologic Engineering Center

Mixed-Population Frequency Analysis

April 1982

Approved for Public Release. Distribution Unlimited.

TD-17

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to the Department of Defense, Executive Services and Communications Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE *(DD-MM-YYYY)*
April 1982

2. REPORT TYPE
Training Document

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
Mixed Population Frequency Analysis

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Edward C. Morris

5d. PROJECT NUMBER

5e. TASK NUMBER

5F. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
US Army Corps of Engineers
Institute for Water Resources
Hydrologic Engineering Center (HEC)
609 Second Street
Davis, CA 95616-4687

8. PERFORMING ORGANIZATION REPORT NUMBER
TD-17

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/ MONITOR'S ACRONYM(S)

11. SPONSOR/ MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This training document provides guidance on the development of frequency curves from annual peak discharges that are segregated into two populations. While the procedures contained in this document use annual peaks caused by hurricane and non-hurricane events, the methods apply equally well to events caused by other phenomena such as rainfall and snowmelt.

15. SUBJECT TERMS
statistical methods, flood frequency, frequency analysis, annual flood

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	UU	52	19b. TELEPHONE NUMBER

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39-18

Mixed-Population Frequency Analysis

April 1982

US Army Corps of Engineers
Institute for Water Resources
Hydrologic Engineering Center
609 Second Street
Davis, CA 95616

(530) 756-1104
(530) 756-8250 FAX
www.hec.usace.army.mil

TD-17

FOREWORD

This training document provides guidance on the development of frequency curves from annual peak discharges that are segregated into two populations. While the procedures contained in this document use annual peaks caused by hurricane and non-hurricane events, the methods apply equally well to events caused by other phenomena such as rainfall and snowmelt.

ACKNOWLEDGEMENTS

The procedures described in this document have been used by the Hydrologic Engineering Center (HEC) in several special assistance studies performed by Messrs. Harold E. Kubik and Leo R. Beard. The Louisville District was responsible for providing the impetus to write this training document by indicating the need for better documentation on the subject. The author wishes to thank Mr. George E. Herbig of the Louisville District, Mr. Doug Speers of the North Pacific Division, and Messrs. Harold E. Kubik and Arlen D. Feldman of the HEC for their review and helpful suggestions.

MIXED POPULATION FREQUENCY ANALYSIS

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

The development of frequency curves at certain locations may require special treatment when the events are caused by different types of hydrologic phenomena and/or the frequency curve exhibits a sudden change in curvature. One example where mixed populations occur is along the Atlantic and Gulf coasts. In these areas events are caused by intense tropical and cyclonic storms, which are referred to as hurricane and non-hurricane events. Another example is in the Sierra Nevada region of California where rainfloods tend to occur November through March and snowmelt floods generally occur April through July. This training document discusses the development of a frequency curve from two or more sets of data (populations) that originate from separate causal factors.

The terminology used in this document is as follows. When the frequency curve is derived from two or more separate frequency curves, each developed from a separate population, the resultant curve is referred to as the combined-population frequency curve. When the resultant frequency curve is derived directly from annual peak data that have not been segregated according to causal factors, it is referred to as a mixed-population frequency curve. This document discusses when and how to develop a combined-population frequency curve from hurricane and non-hurricane populations. The equations can be used to develop a combined-population frequency curve from other mixed populations.

Chapter 2 discusses the merits of mixed-population versus combined-population frequency analyses. The procedure for developing a combined-population frequency curve is described in Chapter 3. Examples of mixed-population and combined-population frequency curves are contained in Chapter 4.

CHAPTER 2. WHEN TO USE A COMBINED-POPULATION FREQUENCY ANALYSIS

The combined-population frequency approach should be considered when the frequency curves derived from mixed populations exhibit rather sudden breaks in the curvature of the frequency curves. Sometimes unusually large or small skew coefficients may be an indication of mixed populations. Unusual regional skew coefficients are generally considered to be greater than 0.7 and less than -0.4.

The sudden break in a frequency curve is often caused by several large events that depart significantly from the trend of the rest of the data. These large events are frequently produced by a different type of hydrologic phenomena; such as hurricanes in a normally rainfall series, rainflood events in a basically snowmelt series (U.S. Geological Survey, 1978), or thunderstorm events in a basically winter rainstorm series. A combined-population analysis is often used to solve this problem, but because of the additional effort required to use this approach, it is not always advantageous to do so. This chapter discusses the considerations involved with making such a decision.

The primary motivation behind a combined-population analysis is to provide a better fit between the analytically derived distribution and the plotting positions than can be obtained with a mixed-population frequency analysis. If the extreme flood events are considered to be the largest in a time period greater than the systematic annual peak flows, then procedures contained in Bulletin 17B, Guidelines for Determining Flood Flow Frequencies, (U.S. Water Resources Council, 1981) may be applicable. The weighting of high events, based on an extended (histroic)period of record, can reduce the departure of the high events from the analytical frequency curve. If historical information is available and the incorporation of these data in a frequency analysis provides a good fit to the plotted data, then a combined-population frequency analysis may not be warranted.

When the historical adjustment does not provide a reasonable fit, or if historical information is not available, then the combined-population frequency approach should be considered. If it is not clear that the one population is responsible for the sudden change in curvature in a fairly large number of cases, then a standard frequency analysis using the mixed-population approach is preferred for three reasons. First, it may be difficult to identify all the events for each population. Second, if there are a small number of occurrences of one population, the resultant frequency curves are not reliable and smoothing of the computed statistics is required. And third, much effort must be expended in deriving generalized skew coefficients for each of the separate populations.

A special consideration for analysis of hurricane and non-hurricane events is the size of the drainage area. In a small drainage area the rainfall intensity of a non-hurricane event can often be equal to that of the hurricane event. As the size of the basin of interest increases, the chance of a non-hurricane event equaling the intensity of a hurricane event decreases. Therefore, the effect of the hurricane events on the small drainage areas is not as pronounced. The Hydrologic Engineering Center has found that catchments less than 500 square miles generally will not require a special hurricane analysis. The drainage area is not a consideration in the decision to segregate rainfall and snowmelt events.

In cases where the sudden departures in curvature are noted in some stations but not in others, the region may need to be subdivided into two separate areas and separate regional analyses employed in each area. Care must be taken to be sure that there are sufficient stations in each area to perform a regional analysis.

Another important consideration is the independence of events. If the data in one of the series is not independent of data in the other series, then a coincident frequency analysis rather than a combined-population frequency analysis will be warranted.

CHAPTER 3. PROCEDURE FOR PERFORMING A COMBINED-POPULATION FREQUENCY ANALYSIS

This chapter describes the procedure for developing a combined-population frequency curve from hurricane and non-hurricane induced events. Section 3.1 discusses the selection of events. Section 3.2 summarizes the procedures and provides references for performing a standard frequency analysis. Section 3.3 describes several methods for determining annual frequency curves from a set of events that do not occur every year. The procedure for combining two frequency curves is reviewed in Section 3.4 and the development of regional relationships used to develop frequency curves at ungaged sites is mentioned in Section 3.5. Procedures to calculate an approximate expected probability adjustment and estimates of confidence limits are contained in Sections 3.6 and 3.7, respectively.

Section 3.1. Data Selection

The first step is to obtain the necessary data to perform the mixed-population frequency analysis. Usually the annual peaks can be obtained directly from the U.S. Geological Survey (USGS) Water Supply Papers. Because of the effort required to gather data for the combined-population frequency analysis, the mixed-population approach should be completed first to determine if additional analyses are warranted.

The collection of data for the combined-population frequency analysis is the next step and involves the determination of the causal factors of the events, and the identification of the largest annual event in each population.

Hurricane events can be identified by studying publications, such as the U.S. Department of Commerce's report on Tropical Cyclones of the North Atlantic Ocean (1965), that specify the hurricane tracks, intensity, and dates of occurrence at selected locations. Next, the dates of these events can be compared with flood events in the USGS Water Supply Papers to determine if the dates of the discharge events correspond to those of the hurricanes. This methodology is approximate because the exact location and the areal extent of the hurricane is not known, and it is difficult to distinguish between a remnant of a hurricane event and a non-hurricane event. Hurricane events do not occur every year over most drainage areas and require special procedures to develop a frequency curve as discussed in Section 3.3.

Another typical application of the combined-population frequency analysis is the division of the year into seasons or months. Bulletin 17B states that "separation by calendar periods in lieu of separation by events is not considered hydrologically reasonable unless the events in the separated periods are clearly caused by different hydrometeorological conditions." The HEC has found that if the data are segregated into too many seasons, then one or more of the seasonal frequency curves may contain one or two large events and many small ones. This causes the seasonal curves to have a very steep slope; and when the seasonal curves are combined into a single annual curve, it causes the upper end of the annual frequency curve to be unreasonably high. In addition, as stated in the U.S. Army Corps of Engineers Engineering Manual on Hydrologic Frequency Analysis (1980), "the combined curve will very likely fit the annual curve only in the middle parts of the curve, and the lower end of the curve will have a partial duration shape as many small events have been included in the analysis."

Section 3.2. Standard Frequency Analysis

If a complete series of annual peaks can be identified, a standard analysis can be performed to determine the annual peak discharge frequency curve. These procedures have been extensively documented in numerous publications (U.S. Army Corps of Engineers, 1962 and 1975a; and U.S. Water Resources Council, 1981), and are briefly summarized in this section. A frequency curve is developed from the annual series of data for each population using the procedures described below.

The first step is to determine graphical plotting positions that define the exceedance probability associated with each discharge. The annual peak data are ranked in descending order and a plotting position is determined using one of several different equations. One of the most common is the Weibull plotting position equation shown below:

$$P = \frac{m}{N+1} \tag{3.1}$$

where P = exceedance probability corresponding to the event of rank m

m = rank of the event

N = number of events

This equation was developed so that the exceedance probability associated with the highest ranked event would be correct, on the average. Equation 3.2 is another commonly used plotting position which is an approximation of the Beard or median plotting position (U.S. Army Corps of Engineers, 1962).

$$P = \frac{m-0.3}{N+0.4} \tag{3.2}$$

The median plotting position was developed so that the exceedance probability associated with the largest event would have an equal chance of being too high or too low. Once the plotting positions have been determined, the exceedance probability and discharge coordinates are plotted on the appropriate probability paper.

An analytical frequency curve is then calculated using the recommended probability distribution. The U.S. Water Resources Council (1981) recommends that the log-Pearson type III distribution, with a weighted skew coefficient, be used to model annual peak discharges. However, WRC's conclusions and generalized skew coefficient map were based on annual peak data that were not segregated according to causal factors. If the log-Pearson type III distribution is desired to model a segregated series, then the investigator will either need to accept the fundamental uncertainty of a calculated skew coefficient, or perform the necessary studies for developing a generalized skew relationship for each type of series. Unless the annual series in a number of stations clearly contain non-zero skew coefficients, a log-normal distribution is recommended.

The analytical frequency curve for each population is calculated and plotted along with the corresponding graphical plotting points. The expected probability adjustment and the confidence limits for the analytical curve are not determined until the combined-population frequency curve has been derived, as described in Sections 3.6 and 3.7.

Section 3.3. Development of Frequency Curves From a Truncated Series

Special frequency analyses are required when events in a series do not occur every year. This section discusses two procedures that have been used to develop hurricane frequency curves in several HEC studies.

The first procedure used at the HEC was developed by Beard (U.S. Army Corps of Engineers, 1958). He recommended that a standard frequency analysis be performed on the hurricane events yielding a curve based on the number of hurricane events per 100 events. This curve is a conditional frequency curve, identical in concept to the one discussed by the U.S. Water Resources Council (1981). The exceedance probabilities of this curve are then multiplied by N_H/N_T, where N_H is the number of hurricane events and N_T is the total number of years of record. While this adjustment can dramatically affect the lower end of the hurricane curve, it causes only a moderate shift of the frequency curve at its upper end. This technique has not been used in many applications because it is usually considered valid when only less than 25 percent of the data is missing. Because this is seldom the case with hurricane events, an alternative procedure has been used by the HEC in several applications (U.S. Army Corps of Engineers, 1965 and 1975b).

The first step in this alternative procedure is to compute plotting positions of the data series using either Equation 3.1 or 3.2 in the same manner as described in Section 3.2 except that N is the number of years rather than the number of events. The frequency curve is then developed by drawing a best-fit line through the plotting positions. This line can be based either by eye or by a modified regression technique which provides a more rigorous mathematical estimate (U.S. Army Corps of Engineers, 1959). However, the modified regression procedure places equal weight on each of the data points, so that one outlier can drastically affect the derived line. The slope of the line developed by one of these procedures is the standard deviation. The mean of the hurricane events is obtained by extending the adopted line and noting the discharge associated with the 0.50 exceedance probability.

Due to the small sample typically used to develop a frequency curve in this manner, there is a great deal of uncertainty in both the mean and standard deviation. This deviation may vary considerably at different geographic locations. Therefore the mean and standard deviation are often plotted versus the drainage area at each gaged site to provide a basis for selecting a regional value. Different mean and standard deviation relationships may be adopted for different river systems (U.S. Army Corps of Engineers, 1975b) or a single relationship for the mean and a single value for the standard deviation may be adopted for the entire region (U.S. Army Corps of Engineers, 1965). A zero skew coefficient is generally adopted unless there is a regional trend that can be rationalized to be caused by known climatic or basin characteristics.

Because of the difficulty in identifying hurricane events, and because of their small sample size, there is a great deal of uncertainty in estimating the mean and standard deviation at individual sites. Therefore, even though the smoothing techniques may be highly subjective, they are desirable.

Section 3.4. Combining Frequency Curves

The procedure for combining frequency curves developed from independent annual series has been widely documented (U.S. Army Corps of Engineers, 1958, 1965, 1975b, and 1980). The general equation for combining multiple frequency curves is:

$$P_c = 1-(1-P_1)(1-P_2)\ldots(1-P_n) = 1-\prod_{i=1}^{n}(1-P_i) \qquad (3.3)$$

where P_c is the exceedance probability of the combined-population frequency curve for the selected discharge

$P_1, P_2, \ldots P_n$ are the exceedance probabilities associated with a selected discharge from frequency curve numbers 1, 2, through n

n is the number of frequency curves that are combined

If only two curves are combined, then Equation 3.3 reduces to:

$$P_c = P_1 + P_2 - P_1 P_2 \qquad (3.4)$$

Equations 3.3 and 3.4 are only valid when each of the frequency curves used to develop a combined curve are assumed to be independent. (See Appendix I for development of these equations.)

Section 3.5. Combined-Population Frequency Curves in Ungaged Areas

This section describes several procedures for developing frequency curves at ungaged sites using combined-population frequency analysis results at gaged sites. The investigator is faced with the choice of determining regional relationships that calculate a mixed-population curve directly or developing regional relationships that calculate a frequency curve for each population. While the latter procedure is more theoretically appealing, there is often greater uncertainty in individual frequency curves developed from the separate populations, and consequently the regional relationships are apt to contain a great deal of uncertainty. Unless the investigator is very confident in the analyses and data used to develop each of the separate-population frequency curves, a mixed-population approach might be warranted because it is simpler to perform and may provide the same level of reliability.

The HEC developed separate regional relationships for hurricane and non-hurricane events on the Kanawha River (U.S. Army Corps of Engineers, 1965). Three empirical curves were developed that described the drainage area versus the mean of the logarithm of the hurricane events, the mean of the logarithm of the non-hurricane events, and the standard deviation of the logarithm of the non-hurricane events. Because there were inadequate data on hurricane floods, a uniform standard deviation of the logarithms of the hurricane events was adopted. These relationships could be developed for hurricane and non-hurricane frequency curves and then be combined using Equation 3.4.

An alternative approach was used in the Tropical Storm Agnes Study (U.S. Army Corps of Engineers, 1975b). In this report, separate regional relationships for calculating hurricane and non-hurricane events were developed for ungaged areas along the major rivers. Regression equations were determined from the mean and standard deviation of the logarithm of the mixed-population frequency curves. These equations, along with a regional skew map, were developed for use in the ungaged areas that were not along the major rivers. By developing a mixed-population curve directly at the ungaged sites, the tremendous uncertainty involved with regionalization of the hurricane events was avoided. However, the use of this approach did not address the problem of the sharply skewed frequency curve.

Section 3.6. Expected Probability

The expected-probability methodology, as proposed by Beard (U.S. Army Corps of Engineers, 1962) adjusts frequency estimates so that the average of the exceedance probabilities for many different sites is closer to the true population exceedance probability. It was developed assuming a normal distribution but has also been used to adjust frequency curves developed using a log-Pearson type III distribution. Hardison and Jennings (1972) have shown that this adjustment reduces the bias in samples drawn from a log-Pearson type III distribution with a known skew coefficient. Lloyd (1978) indicated that similar estimating procedures could be developed for other distributions. However, he also indicated that their mathematical derivation is extremely complex.

The underlying distribution of the combined-population frequency curve developed using Equation 3.4 is unknown. Therefore the correct adjustment to cause the average of the probabilities from many sample frequency curves to equal the "true" population exceedance probability is also unknown. The trade-off becomes one of whether a possibly incorrect mathematical adjustment should be applied, or no adjustment at all. The HEC has made the adjustment in their studies because even though the adjustment may not be known with much uncertainty, it was felt that it would cause less bias in frequency estimates than no adjustment.

The expected probability adjustment can be made using Chart 40 in "Statistical Methods in Hydrology" (U.S. Army Corps of Engineers, 1962) or by employing Equation 11-1 in the Water Resource Council Guidelines (1981) with a cumulative distribution of the student's t distribution (Benjamin and Cornell, 1970). Both procedures for making the expected probability adjustment require using the number of events, n, in the sample. For a combined-population frequency curve at a gaged site, the HEC has defined n as the larger of the number of events used to develop the non-hurricane or the hurricane frequency curve. However, at ungaged sites, the mean and standard deviation are often developed from separate regional equations. The worth of the mean and standard deviation will depend upon the correlation coefficient and the number of years of record at the gaged sites used to develop these relationships. The HEC has been unable to develop a theoretically appealing mathematical relationship to establish an equivalent period of record at an ungaged site. Typically, the average number of years of record at the gaged sites in the region has been calculated and used to make the expected probability adjustment at ungaged locations.

Section 3.7. Confidence Limits for Combined-Population Frequency Curves

Confidence limits for the normal distribution can be calculated using a non-central t distribution (Resnikoff and Lieberman, 1957). The non-central t has also been employed to calculate confidence limits for a log-Pearson type III distribution with a skew coefficient between ±0.5 (U.S. Water Resources Council, 1981). However, in the latter case, only the uncertainties in the mean and standard deviation are accounted for. When the underlying distribution is not normal, the calculation of confidence limits based on the non-central t may not be theoretically valid.

Establishing confidence limits for a combined-population frequency curve has the same theoretical difficulties as making the expected probability adjustment: either calculate a possibly incorrect mathematical confidence limit or none at all. The HEC has used two alternate procedures to calculate approximate confidence limits for the combined-population frequency curve as described in this document. These techniques are only suggested solutions to the problem at this time. Further analysis needs to be carried out to verify these procedures or develop new, more reliable procedures.

Method 1

The first procedure for calculating confidence limits employs Exhibit 6 in "Statistical Methods in Hydrology" (U.S. Army Corps of Engineers, 1962). This chart tabulates an error of the estimated value, $E_{N,P,c}$, which is a function of the years of record N, the exceedance probability P, and the confidence level c. The confidence limits can then be calculated using the equation:

$$X_{P,c} = X_P + E_{N,P,c} * S \quad (3.5)$$

where S is the standard deviation of the frequency curve, X_P is the logarithm of the discharge at the exceedance probability P, and $X_{P,c}$ is the logarithm of the confidence limit with exceedance probability P. The tabulated values of $E_{N,P,c}$ are positive for the upper confidence limits and negative for the lower limits. If values of $E_{N,P,c}$ are needed which are not tabulated in Exhibit 6 (U.S. Army Corps of Engineers, 1962), they can be calculated by the following equations:

$$E_{N,P,c} = x_{P,c} * \sqrt{\frac{N-1}{N}} - K_P \quad (3.6)$$

$$x_{P,c} = \frac{t_{N,P,c}}{\sqrt{N-1}} \quad (3.7)$$

where K_P is the normal deviate for exceedance probability P, and $t_{N,P,c}$ is the non-central t value. Values of $x_{P,c}$, a non-central t argument, for various degrees of freedom (N-1 in this application) can be found in Resnikoff and Lieberman (1957), and K_P is found in numerous sources (e.g., U.S. Water Resources Council, 1981). If non-central t tables are not available, Equations 9-4 through 9-6 in the Water Resources Guidelines can be used to determine approximate values of $K_{P,c}$.

These are related to the error of the estimated value as shown below for the upper limit:

$$E^{U}_{N,P,c} = K^{U}_{P,c} - K_{P} \tag{3.8}$$

and for the lower limit:

$$E^{L}_{N,P,c} = K^{L}_{P,c} - K_{P} \tag{3.9}$$

From Equations 3.6, 3.7, 3.8, and 3.9 it becomes apparent that:

$$K_{P,c} = t_{N,P,c}/\sqrt{N} \tag{3.10}$$

Although the WRC equations 9-4 through 9-6 are only approximate, the HEC has found they are generally satisfactory for the 5% confidence level.

As shown in Equation 3.5, confidence limits for combined-population frequency curve require a value for the standard deviation. Typically the lower end of the curve, which follows the non-hurricane curve, will have a low standard deviation and the upper end, which follows the hurricane curve, will have a high standard deviation. Therefore a procedure is necessary which accounts for this change.

The HEC has employed Equation 3.5 to calculate confidence limits directly for the combined-population frequency curve in past studies (U.S. Army Corps of Engineers, 1975b). A weighted standard deviation is used based on the following equation:

$$S_C = \left(\frac{P_{H,Q}}{P_{H,Q} + P_{N,Q}}\right) S_H + \left(\frac{P_{N,Q}}{P_{H,Q} + P_{N,Q}}\right) S_N \tag{3.12}$$

where S_C is the standard deviation of the combined-population frequency curve for the discharge Q; $P_{H,Q}$ is the exceedance probability from the hurricane frequency curve associated with the discharge Q; $P_{N,Q}$ is the exceedance probability from the non-hurricane frequency curve associated with discharge Q; S_H is the standard deviation of the hurricane frequency curve, and S_N is the standard deviation of the non-hurricane frequency curve. As shown in the example in Chapter 4, this procedure will not always yield valid confidence limits. If there is a large difference between the hurricane and non-hurricane standard deviations and skew coefficients, this can lead to an irregularly shaped lower confidence limit below the intersection of the two curves.

Method 2

An alternate procedure is to calculate separate confidence limits for the hurricane and non-hurricane frequency curves. These limits can be combined using equation 3.4. While this method avoids the problems of the irregularly shaped confidence limits, it does yield confidence bounds that are perhaps too close to the combined-population frequency curve.

If the standard deviation and skew coefficients are reasonably close, the method using Equations 3.5 -3.12 should provide satisfactory results. When this is not true the second method that uses Equation 3.4 to combine the separately derived confidence limits may provide more consistent results. In either case the investigator must carefully examine the derived confidence limits to be sure they appear reasonable and remember that both procedures provide only rough estimates of the actual uncertainty in the combined-population frequency curve.

CHAPTER 4. EXAMPLES OF MIXED-POPULATION AND COMBINED-POPULATION FREQUENCY ANALYSIS

This chapter contains examples of mixed-population and combined-population frequency analysis for a stream gaging station on West Conewago Creek near Manchester, Pennsylvania. Section 4.1 summarizes the data and Sections 4.2 and 4.3 describe the computations for mixed and combined analyses, respectively. The frequency curves derived in these Chapters are for illustrative purposes only; they are not to be construed as the recommended curves for West Conewago Creek.

Section 4.1. Data Selection

Figure 4.1 contains an excerpt from a U.S.G.S. water supply paper and a list of additional annual peaks for West Conewago Creek near Manchester, PA for the years 1929 to 1972. A separate analysis has indicated that five hurricane events were interspersed throughout the 44 years of recorded data. Both the hurricane and non-hurricane annual peak discharges for these years are shown in Figure 4.2.

5740. West Conewago Creek near Manchester, Pa.
(Published as "Conewago Creek" prior to 1932)

Location.--Lat 40°04'55", long 76°43'10", 500 ft upstream from bridge on State Highway 24, 0.7 mile downstream from Little Conewago Creek, and 1.5 miles north of Manchester, York County.

Drainage area.--510 sq mi.

Gage.--Recording. Datum of gage is 263.68 ft above mean sea level, datum of 1929.

Stage-discharge relation.--Defined by current-meter measurements.

Bankfull stage.--7 ft.

Remarks.--Base for partial-duration series, 10,800 cfs.

Peak stages and discharges of West Conewago Creek near Manchester, Pa.

Water year	Date	Gage height (feet)	Discharge (cfs)
1929	Mar. 6, 1929	11.70	11,800
	Apr. 17, 1929	15.31	20,300
	May 3, 1929	13.79	16,500
1930	Oct. 2, 1929	12.58	13,700*
	Mar. 8, 1930	11.16	10,800
1931	Apr. 2, 1931	9.14	6,850
1932	Mar. 28, 1932	12.12	11,900
1933	Oct. 19, 1932	13.37	14,200
	Nov. 1, 1932	12.62	12,800
	Apr. 20, 1933	13.26	14,100
	Aug. 24, 1933	24.14	47,600*
1934	Sept. 15, 1934	13.71	15,300
	Sept. 17, 1934	17.41	24,900
	Sept. 30, 1934	17.20	24,400
1935	Dec. 1, 1934	15.86	20,700
1936	Mar. 12, 1936	13.02	13,700
	Mar. 19, 1936	a17.08	-
	Apr. 6, 1936	12.93	13,500
	June 13, 1936	11.93	11,400
1937	Feb. 22, 1937	11.73	12,100
	Apr. 27, 1937	12.08	12,900
1938	Oct. 23, 1937	11.27	11,200
	Nov. 13, 1937	13.82	16,800
1939	Feb. 4, 1939	13.70	16,500
	Mar. 1, 1939	11.18	11,000
1940	Apr. 9, 1940	12.18	12,500
	Apr. 20, 1940	15.85	21,300
	Sept. 1, 1940	11.63	11,200*
1941	Apr. 6, 1941	11.16	10,400
1942	May 22, 1942	12.79	13,800
	June 5, 1942	12.58	13,400
	Aug. 18, 1942	14.28	17,400
1943	Dec. 30, 1942	14.09	16,900
1944	Nov. 9, 1943	17.33	25,500
	Mar. 13, 1944	11.36	10,800
	Mar. 24, 1944	11.39	10,800
	May 7, 1944	11.37	10,800
1945	July 19, 1945	11.39	11,100
1946	Nov. 29, 1945	15.86	21,600
	June 2, 1946	15.74	21,000
1947	May 22, 1947	13.66	16,000
1948	Jan. 2, 1948	10.62	9,980
1949	Dec. 30, 1948	13.74	16,000
	Jan. 6, 1949	12.96	14,400
1950	Mar. 23, 1950	12.51	13,300
	May 19, 1950	12.18	12,700
1951	Nov. 26, 1950	12.53	13,300
	Dec. 4, 1950	13.91	16,400
	Feb. 7, 1951	11.67	11,700
	Feb. 21, 1951	11.64	11,500
1952	Feb. 4, 1952	11.93	12,100
	Mar. 11, 1952	13.88	16,400
	Apr. 28, 1952	11.72	11,700
1953	Nov. 22, 1952	13.96	16,700
	Jan. 24, 1953	12.66	13,700
1954	Mar. 2, 1954	8.30	5,740
1955	Mar. 22, 1955	14.10	16,900
	Aug. 13, 1955	11.29	10,900*
1956	Oct. 14, 1955	12.92	14,200
1957	Dec. 15, 1956	11.37	11,100
	Apr. 6, 1957	11.34	10,900
1958	Dec. 21, 1957	13.77	16,200
	Dec. 27, 1957	13.14	14,600
	Feb. 28, 1958	12.97	14,400
	Mar. 26, 1958	12.03	12,300
	May 6, 1958	11.74	11,700
1959	Jan. 22, 1959	10.07	8,720
1960	Apr. 4, 1960	12.55	13,500
1961	Feb. 26, 1961	13.00	14,400
	Apr. 13, 1961	12.84	14,000

a Backwater from ice.

1962	**11000**	**1968**	**16400**
1963	**15000**	**1969**	**15500**
1964	**11500**	**1970**	**21300**
1965	**18100**	**1971**	**15700**
1966	**16000**	**1972**	**81700***
1967	**19000**		

*** Hurricane events**

Figure 4.1 Annual peak discharges at West Conewago Creek
(U.S. Geological Survey, 1968)

Year	Non-hurricane Discharge (cfs)	Hurricane Discharge (cfs)
1930	10800	13700
1933	14100	47600
1940	21300	11200
1955	16900	10900
1972	12700	81700

Figure 4.2 Hurricane and non-hurricane annual peak discharges

Section 4.2. Mixed-Population Frequency Analysis

The annual peaks for West Conewago Creek, irrespective of causal factors, were provided as input to the Flood Flow Frequency Computer Program (U.S. Army Corps of Engineers, 1976) which was used to perform a frequency analysis. A generalized skew coefficient of 0.5 was obtained from the map in Bulletin 17B (U.S. Water Resources Council, 1981). The plotting positions are shown in Figure 4.3, the final statistics and frequency ordinates in Figure 4.4, and the frequency curve in Figure 4.5. It is evident that the two largest discharges, which are both hurricane events, caused a very high calculated skew coefficient. Both these events depart significantly from the analytical curve. Either historical information should be sought to determine if, in fact, the hurricane events were the largest over a historical period greater than the 44 years of recorded data, or a combined-population analysis should be performed.

Because there was no readily accessible historical information at West Conewago Creek, information at adjacent gages was sought. The closest long term stream was located on the Susquehanna River hear Harrisburg, Pennsylvania, and had a continuous record since 1889.

```
-ANNUAL PEAKS   - WEST CONEWAGO CREEK NR MANCHESTER,PA
*********************************************************************
*......DATA ANALYZED.......*.............ORDERED DATA...........*
*                           *          WATER              WEIBULL  *
* MON DAY  YEAR      FLOW   *   RANK  YEAR       FLOW    PLOT POS *
*---------------------------*-------------------------------------*
*  -0  -0  1929   20300.   *     1   1972     81700.     .0222   *
*  -0  -0  1930   13700.   *     2   1933     47600.     .0444   *
*  -0  -0  1931    6850.   *     3   1944     25500.     .0667   *
*  -0  -0  1932   11900.   *     4   1934     24900.     .0889   *
*  -0  -0  1933   47600.   *     5   1946     21600.     .1111   *
*  -0  -0  1934   24900.   *     6   1940     21300.     .1333   *
*  -0  -0  1935   20700.   *     7   1970     21300.     .1556   *
*  -0  -0  1936   13700.   *     8   1935     20700.     .1778   *
*  -0  -0  1937   12900.   *     9   1929     20300.     .2000   *
*  -0  -0  1938   16800.   *    10   1967     19000.     .2222   *
*  -0  -0  1939   16500.   *    11   1965     18100.     .2444   *
*  -0  -0  1940   21300.   *    12   1942     17400.     .2667   *
*  -0  -0  1941   10400.   *    13   1943     16900.     .2889   *
*  -0  -0  1942   17400.   *    14   1955     16900.     .3111   *
*  -0  -0  1943   16900.   *    15   1938     16800.     .3333   *
*  -0  -0  1944   25500.   *    16   1953     16700.     .3556   *
*  -0  -0  1945   11100.   *    17   1939     16500.     .3778   *
*  -0  -0  1946   21600.   *    18   1951     16400.     .4000   *
*  -0  -0  1947   16000.   *    19   1952     16400.     .4222   *
*  -0  -0  1948    9980.   *    20   1968     16400.     .4444   *
*  -0  -0  1949   16000.   *    21   1958     16200.     .4667   *
*  -0  -0  1950   13300.   *    22   1947     16000.     .4889   *
*  -0  -0  1951   16400.   *    23   1949     16000.     .5111   *
*  -0  -0  1952   16400.   *    24   1966     16000.     .5333   *
*  -0  -0  1953   16700.   *    25   1971     15700.     .5556   *
*  -0  -0  1954    5740.   *    26   1969     15500.     .5778   *
*  -0  -0  1955   16900.   *    27   1963     15000.     .6000   *
*  -0  -0  1956   14200.   *    28   1961     14400.     .6222   *
*  -0  -0  1957   11100.   *    29   1956     14200.     .6444   *
*  -0  -0  1958   16200.   *    30   1930     13700.     .6667   *
*  -0  -0  1959    8720.   *    31   1936     13700.     .6889   *
*  -0  -0  1960   13500.   *    32   1960     13500.     .7111   *
*  -0  -0  1961   14400.   *    33   1950     13300.     .7333   *
*  -0  -0  1962   11100.   *    34   1937     12900.     .7556   *
*  -0  -0  1963   15000.   *    35   1932     11900.     .7778   *
*  -0  -0  1964   11500.   *    36   1964     11500.     .8000   *
*  -0  -0  1965   18100.   *    37   1945     11100.     .8222   *
*  -0  -0  1966   16000.   *    38   1957     11100.     .8444   *
*  -0  -0  1967   19000.   *    39   1962     11100.     .8667   *
*  -0  -0  1968   16400.   *    40   1941     10400.     .8889   *
*  -0  -0  1969   15500.   *    41   1948      9980.     .9111   *
*  -0  -0  1970   21300.   *    42   1959      8720.     .9333   *
*  -0  -0  1971   15700.   *    43   1931      6850.     .9556   *
*  -0  -0  1972   81700.   *    44   1954      5740.     .9778   *
*********************************************************************
```

Figure 4.3 Mixed frequency analysis for systematic data at West Conewago Creek - plotting positions

<u>Illustrative Example</u>

FINAL RESULTS

-FREQUENCY CURVE- WEST CONEWAGO CREEK NR MANCHESTER,PA

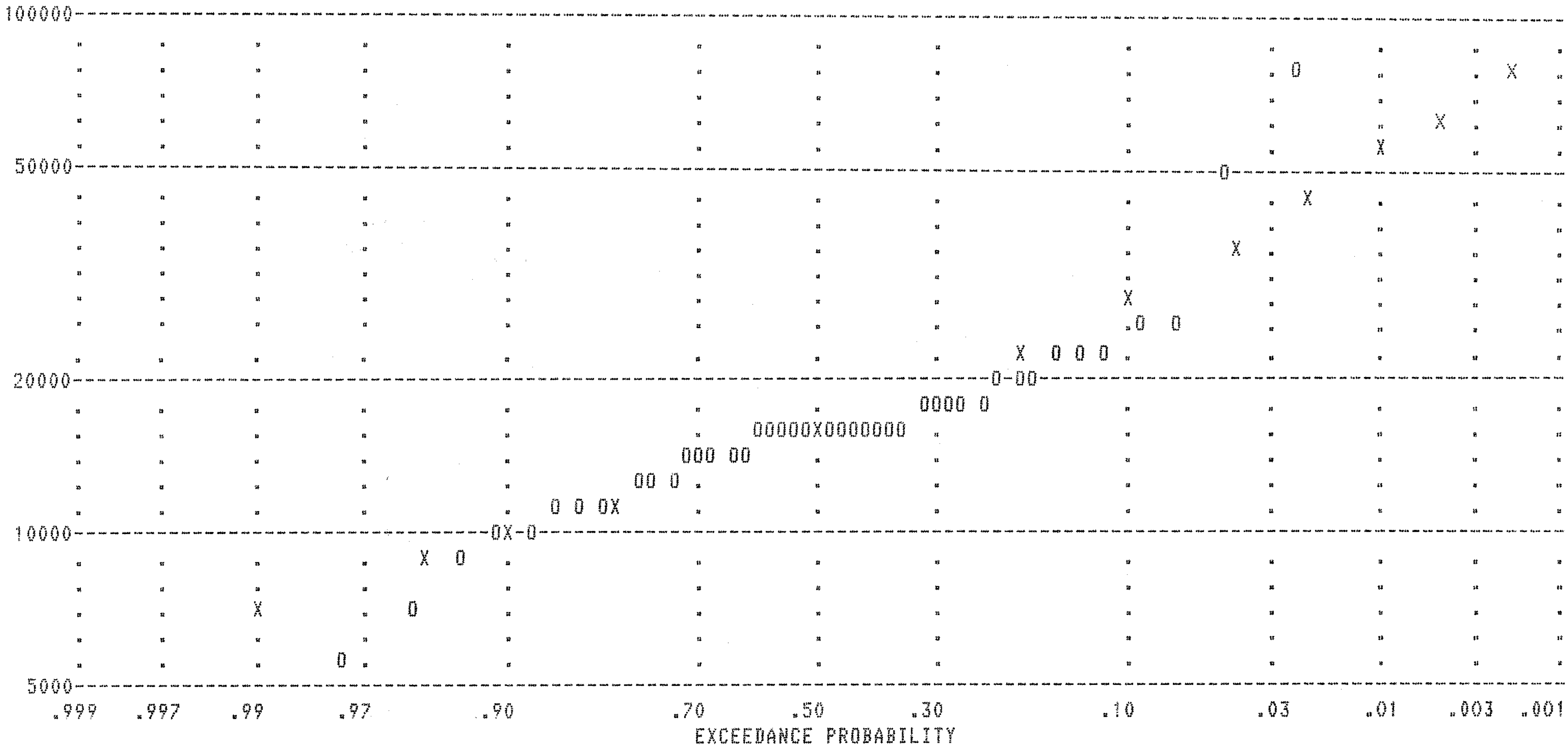
PEAK FLOWS			CONFIDENCE LIMITS	
COMPUTED	EXPECTED PROBABILITY	EXCEEDANCE PROBABILITY	.05 LIMIT	.95 LIMIT
79000.	91700.	.002	114000.	60800.
63500.	70700.	.005	87600.	50300.
53400.	57900.	.010	71300.	43300.
44600.	47300.	.020	57500.	37100.
36900.	38400.	.040	45800.	31400.
28100.	28600.	.100	33300.	24600.
22200.	22400.	.200	25400.	19800.
15000.	15000.	.500	16700.	13400.
10900.	10800.	.800	12200.	9470.
9460.	9370.	.900	10700.	8060.
8530.	8400.	.950	9780.	7150.
7230.	7050.	.990	8430.	5890.

FREQUENCY CURVE STATISTICS		STATISTICS BASED ON	
MEAN LOGARITHM	4.1979	HISTORIC EVENTS	0
STANDARD DEVIATION	.1876	HIGH OUTLIERS	0
COMPUTED SKEW	1.1903	LOW OUTLIERS	0
GENERALIZED SKEW	.5000	ZERO OR MISSING	0
ADOPTED SKEW	.7000	SYSTEMATIC YEARS	44
		TOTAL PERIOD, YEARS	44

Figure 4.4 Mixed frequency analysis for systematic data of West Conewago Creek - statistics and discharges for selected exceedance probabilities.

__Illustrative Example__

FINAL RESULTS
-FREQUENCY PLOT - WEST CONEWAGO CREEK NR MANCHESTER,PA
BASED ON COMPUTED VALUES, FLOW IN CUBIC FEET PER SECOND



LEGEND - O=OBSERVED VALUE, H=HIGH OUTLIER OR HISTORIC VALUE, L=LOW OUTLIER, Z=ZERO OR MISSING X=COMPUTED CURVE

Figure 4.5 Mixed frequency analysis for systematic data at West Conewago Creek - plot of frequency curve

<u>Illustrative Example</u>

The 1972 Agnes event was the largest in the period of record. It was assumed that the 1972 Agnes event at West Conewago Creek was also the largest since 1889. The Flood Flow Frequency computer program was run using this historical information and the resultant statistics and frequency curve are shown in Figures 4.6 and 4.7. In this particular case, the historical adjustment lowers the upper end of the frequency roughly ten percent.

The plotting positions of the two highest events are still well above the analytical frequency curve. The following section illustrates the use of a combined-population frequency analysis to handle this problem.

Section 4.3. Combined-Population Frequency Analysis

This section illustrates the basic steps used to develop a combined-population frequency curve using the data shown in Figures 4.1 and 4.2.

Section 4.3.1. Development of a Non-Hurricane Frequency Curve

The non-hurricane frequency curve was derived from the 44 years of non-hurricane events using the Flood Flow Frequency computer program. The plotting positions, statistics and discharges for selected exceedance probabilities, and the frequency curve are shown in Figures 4.8, 4.9, and 4.10, respectively.

Section 4.3.2. Development of a Hurricane Frequency Curve

Only five hurricane events were noted during the 44 years of recorded flows, therefore the hurricane frequency curve will be based on a truncated series. The frequency curve is developed using the second procedure discussed in Section 3.3. Plotting point positions, tabulated in Figure 4.11, were

FINAL RESULTS

-FREQUENCY CURVE- WEST CONEWAGO CREEK NR MANCHESTER,PA

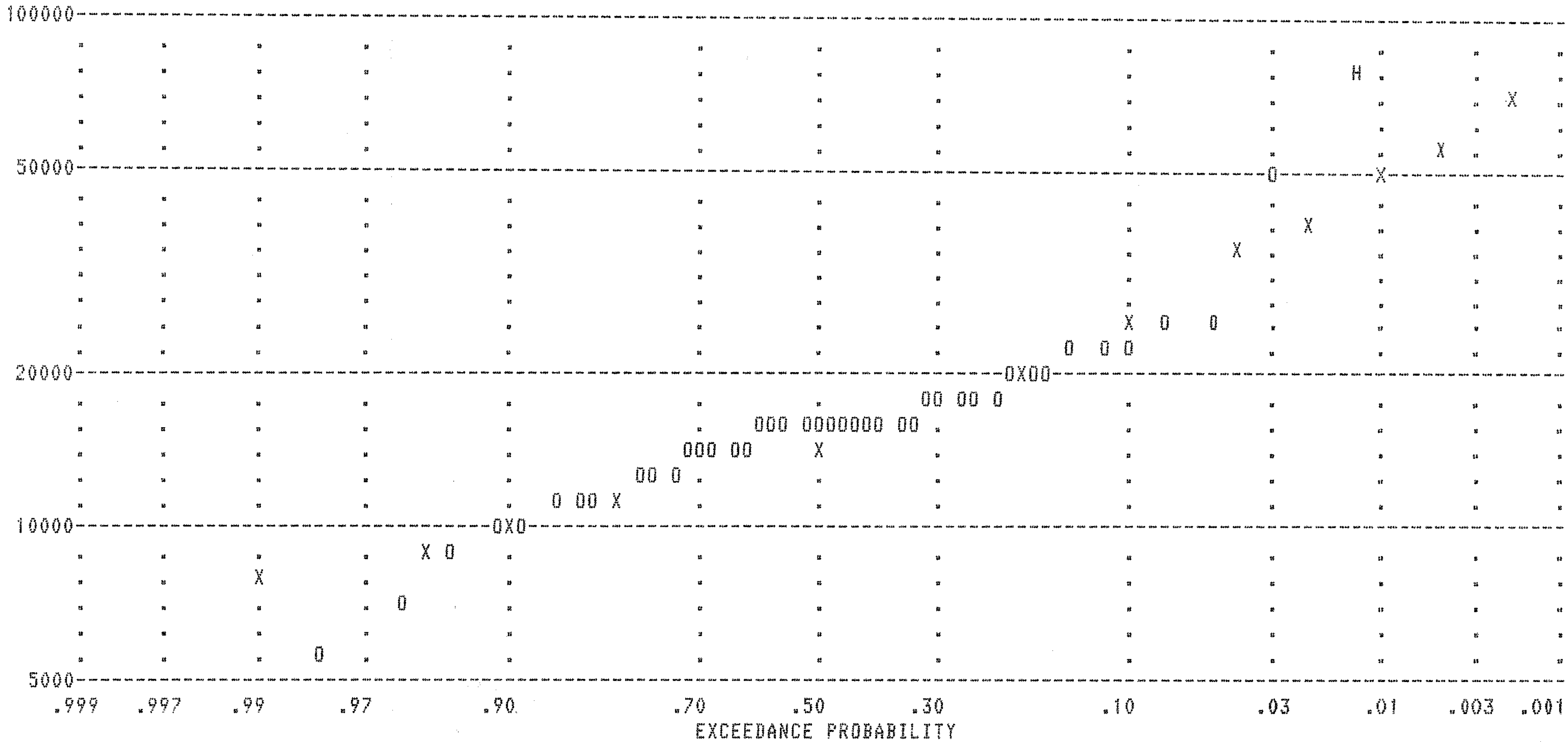
PEAK FLOWS COMPUTED	PEAK FLOWS EXPECTED PROBABILITY	EXCEEDANCE PROBABILITY	CONFIDENCE LIMITS .05 LIMIT	CONFIDENCE LIMITS .95 LIMIT
70800.	81900.	.002	100000.	55400.
57300.	63600.	.005	77500.	46100.
48500.	52500.	.010	63500.	39900.
40800.	43200.	.020	51600.	34400.
34000.	35300.	.040	41600.	29300.
26300.	26800.	.100	30700.	23200.
21100.	21300.	.200	23900.	19000.
14700.	14700.	.500	16200.	13300.
11000.	11000.	.800	12300.	9720.
9770.	9690.	.900	11000.	8450.
8950.	8840.	.950	10100.	7630.
7810.	7660.	.990	8970.	6510.

FREQUENCY CURVE STATISTICS		STATISTICS BASED ON	
MEAN LOGARITHM	4.1900	HISTORIC EVENTS	0
STANDARD DEVIATION	.1715	HIGH OUTLIERS	1
COMPUTED SKEW	.8360	LOW OUTLIERS	0
GENERALIZED SKEW	.5000	ZERO OR MISSING	0
ADOPTED SKEW	.8000	SYSTEMATIC YEARS	44
		TOTAL PERIOD, YEARS	84

Figure 4.6 Mixed frequency analysis for systematic and historical information at West Conewago Creek - statistics and discharges for selected exceedance probabilities.

Illustrative Example

FINAL RESULTS
-FREQUENCY PLOT - WEST CONEWAGO CREEK NR MANCHESTER,PA
BASED ON COMPUTED VALUES, FLOW IN CUBIC FEET PER SECOND



LEGEND - O=OBSERVED VALUE, H=HIGH OUTLIER OR HISTORIC VALUE, L=LOW OUTLIER, Z=ZERO OR MISSING X=COMPUTED CURVE

Figure 4.7 Mixed frequency analysis for systematic and historical information at West Conewago Creek - plot of frequency curve.

Illustrative Example

FINAL RESULTS

-ANNUAL PEAKS - WEST CONEWAGO CREEK NR MANCHESTER,PA

DATA ANALYZED				ORDERED DATA			
MON	DAY	YEAR	FLOW	RANK	WATER YEAR	FLOW	WEIBULL PLOT POS
-0	-0	1929	20300.	1	1944	25500.	.0222
-0	-0	1930	10800.	2	1934	24900.	.0444
-0	-0	1931	6850.	3	1946	21600.	.0667
-0	-0	1932	11900.	4	1940	21300.	.0889
-0	-0	1933	14100.	5	1970	21300.	.1111
-0	-0	1934	24900.	6	1935	20700.	.1333
-0	-0	1935	20700.	7	1929	20300.	.1556
-0	-0	1936	13700.	8	1967	19000.	.1778
-0	-0	1937	12900.	9	1965	18100.	.2000
-0	-0	1938	16800.	10	1942	17400.	.2222
-0	-0	1939	16500.	11	1943	16900.	.2444
-0	-0	1940	21300.	12	1955	16900.	.2667
-0	-0	1941	10400.	13	1938	16800.	.2889
-0	-0	1942	17400.	14	1953	16700.	.3111
-0	-0	1943	16900.	15	1939	16500.	.3333
-0	-0	1944	25500.	16	1951	16400.	.3556
-0	-0	1945	11100.	17	1952	16400.	.3778
-0	-0	1946	21600.	18	1968	16400.	.4000
-0	-0	1947	16000.	19	1958	16200.	.4222
-0	-0	1948	9980.	20	1949	16000.	.4444
-0	-0	1949	16000.	21	1947	16000.	.4667
-0	-0	1950	13300.	22	1966	16000.	.4889
-0	-0	1951	16400.	23	1971	15700.	.5111
-0	-0	1952	16400.	24	1969	15500.	.5333
-0	-0	1953	16700.	25	1963	15000.	.5556
-0	-0	1954	5740.	26	1961	14400.	.5778
-0	-0	1955	16900.	27	1956	14200.	.6000
-0	-0	1956	14200.	28	1933	14100.	.6222
-0	-0	1957	11100.	29	1936	13700.	.6444
-0	-0	1958	16200.	30	1960	13500.	.6667
-0	-0	1959	8720.	31	1950	13300.	.6889
-0	-0	1960	13500.	32	1937	12900.	.7111
-0	-0	1961	14400.	33	1972	12700.	.7333
-0	-0	1962	11100.	34	1932	11900.	.7556
-0	-0	1963	15000.	35	1964	11500.	.7778
-0	-0	1964	11500.	36	1957	11100.	.8000
-0	-0	1965	18100.	37	1945	11100.	.8222
-0	-0	1966	16000.	38	1962	11100.	.8444
-0	-0	1967	19000.	39	1930	10800.	.8667
-0	-0	1968	16400.	40	1941	10400.	.8889
-0	-0	1969	15500.	41	1948	9980.	.9111
-0	-0	1970	21300.	42	1959	8720.	.9333
-0	-0	1971	15700.	43	1931	6850.	.9556
-0	-0	1972	12700.	44	1954	5740.	.9778

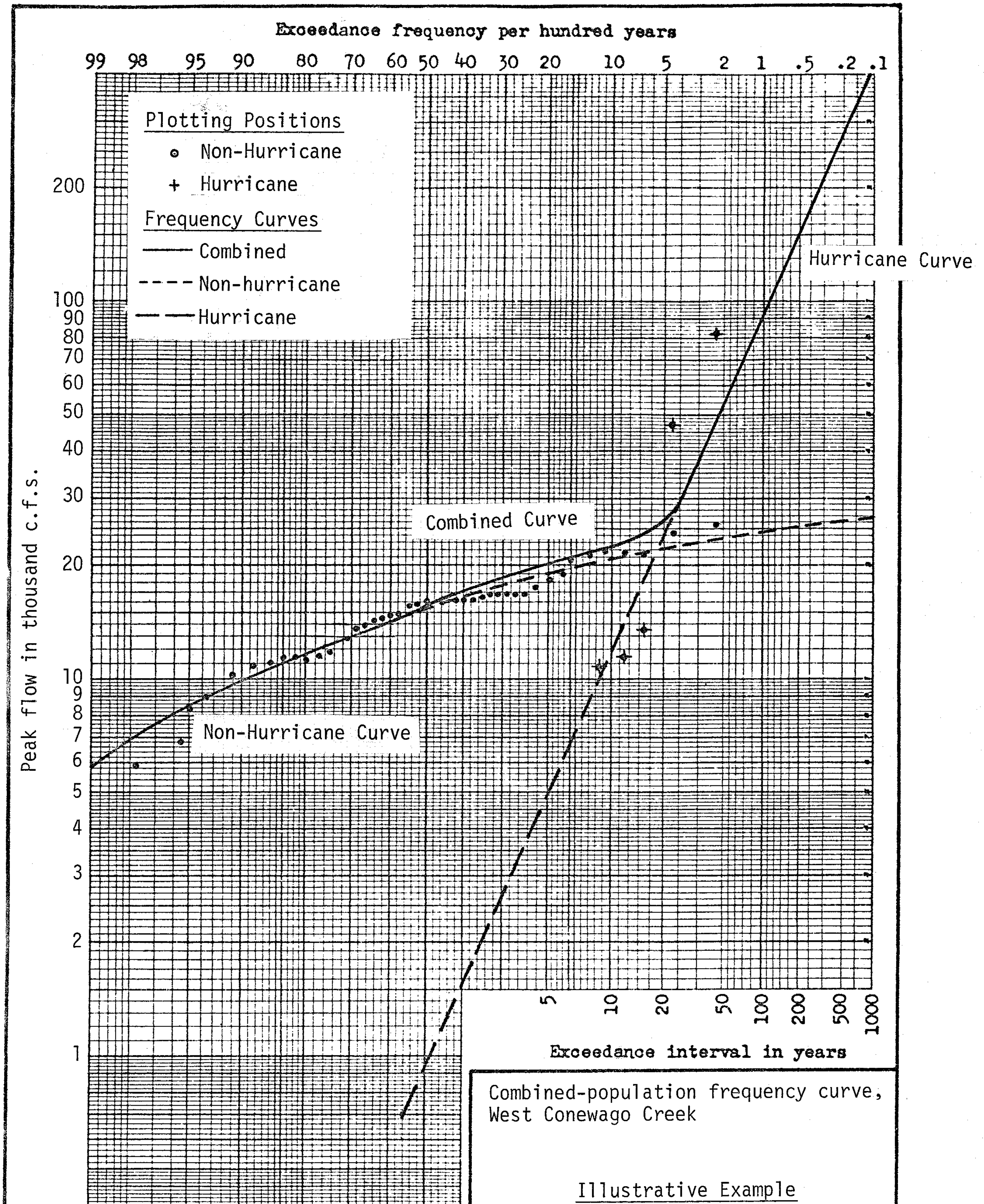
Figure 4.8 Non-hurricane frequency analysis at West Conewago Creek - plotting positions.

<u>Illustrative Example</u>

```
-FREQUENCY CURVE- WEST CONEWAGO CREEK NR MANCHESTER,PA
*****************************************************************
*.......PEAK FLOWS.......*              *...CONFIDENCE LIMITS...*
*            EXPECTED    * EXCEEDANCE   *                       *
*  COMPUTED  PROBABILITY * PROBABILITY  *  .05 LIMIT  .95 LIMIT *
*------------------------*--------------*-----------------------*
*    26600.      27100.  *     .002     *    31000.     23700.  *
*    25700.      26100.  *     .005     *    29700.     23000.  *
*    24900.      25200.  *     .010     *    28600.     22300.  *
*    23900.      24200.  *     .020     *    27400.     21600.  *
*    22800.      23000.  *     .040     *    25900.     20700.  *
*    20900.      21000.  *     .100     *    23400.     19100.  *
*    19000.      19100.  *     .200     *    21000.     17500.  *
*    15200.      15200.  *     .500     *    16500.     14100.  *
*    11500.      11400.  *     .800     *    12500.     10500.  *
*     9710.       9570.  *     .900     *    10700.      8610.  *
*     8330.       8130.  *     .950     *     9300.      7190.  *
*     6030.       5680.  *     .990     *     7020.      4900.  *
*+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++*
*   FREQUENCY CURVE STATISTICS    *     STATISTICS BASED ON     *
*---------------------------------*-----------------------------*
*  MEAN LOGARITHM         4.1651  *  HISTORIC EVENTS        0   *
*  STANDARD DEVIATION      .1330  *  HIGH OUTLIERS      0       *
*  COMPUTED SKEW         -.8398   *  LOW OUTLIERS       0       *
*  GENERALIZED SKEW    -99.0000   *  ZERO OR MISSING    0       *
*  ADOPTED SKEW          -.8000   *  SYSTEMATIC YEARS      44   *
*                                 *  TOTAL PERIOD, YEARS   44   *
*****************************************************************
```

Figure 4.9 Non-hurricane frequency analysis at West Conewago Creek - statistics and discharges for selected exceedance probabilities

Illustrative Example



Combined-population frequency curve, West Conewago Creek

Illustrative Example

Figure 4.10

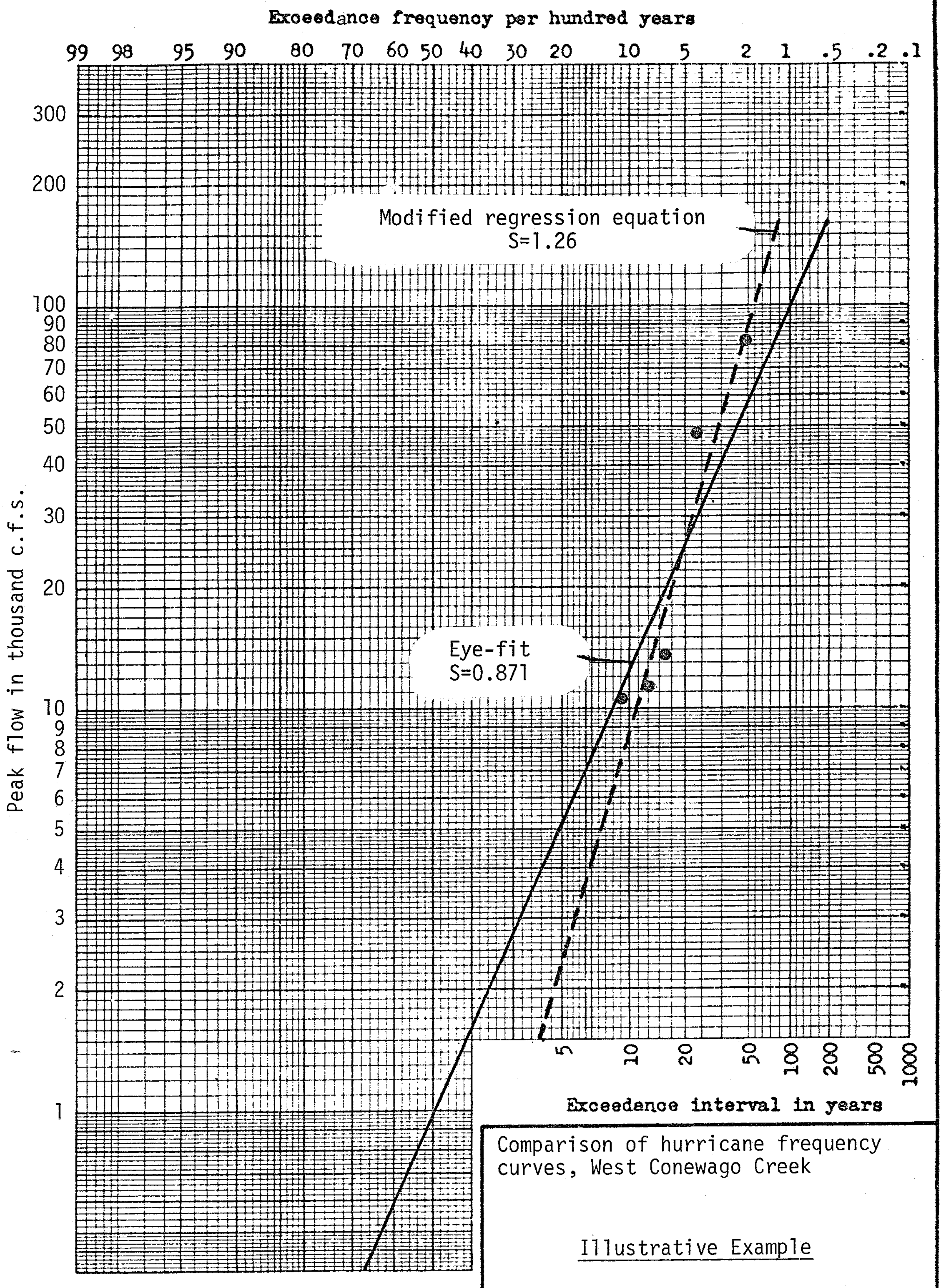
Year	Event	Plotting Position
1972	81700	0.0222
1933	47600	0.0444
1930	13700	0.0667
1940	11200	0.0889
1955	10900	0.1111

Figure 4.11 Hurricane plotting positions

calculated using Equation 3.1. N was defined as 44, the number of years of record. These plotting points are drawn on Figure 4.12 and an eye-fit curve was drawn. In addition, a modified regression line was calculated using a procedure recommended by Beard (U.S. Army Corps of Engineers, 1959). The calculations are shown in Figure 4.13 and the calculated regression line is also drawn on Figure 4.12. The eye-fit curve was selected for inclusion in the combined-frequency curve analysis. The modified regression line was judged to be too steep because of the following. First, it is quite possible that the hurricane event is the largest in a period longer than the 44 recorded years of data as discussed in Section 4.2. Second, the modified regression line is quite sensitive to the magnitudes of the smaller hurricane events. Third, other regional studies have shown that for basins of similar size, the standard deviation varies from 0.6 to 1.2, and the slope of the eye-fit line falls comfortably in the middle of this range. The adopted hurricane frequency curve with a mean of 2.9731 and a standard deviation of 0.871 is displayed with the non-hurricane curve in Figure 4.10.

Section 4.3.3 Development of a Combined-Population Frequency Curve

The hurricane and non-hurricane frequency curves that were developed in the previous two sections are combined using Equation 3.4. Discharges are selected



Comparison of hurricane frequency curves, West Conewago Creek

Illustrative Example

Figure 4.12

I. Tabulation of log Q versus the corresponding K deviate.

Event Q cfs	log Q (X)	K-normal deviate corresponding to plotting position
81700	4.9122	2.0124
47600	4.6776	1.7041
13700	4.1367	1.5242
11200	4.0492	1.3629
10900	4.0374	1.2327

For this example X equals log Q, N is the number of events, and K can be obtained from tables of the standardized normal distribution (Benjamin and Cornell, 1970).

II. Calculation of the standard deviation of the frequency curve.

$$S = \sqrt{\frac{\Sigma X^2 - (\Sigma X)^2/n}{\Sigma K^2 - (\Sigma K)^2/n}} = \frac{S_X}{S_K} = \frac{0.40506}{0.30513} = 1.327$$

III. Calculation of the mean of the frequency curve.

$$M_X = \Sigma X/N = 4.3626$$

$$M_K = \Sigma K/N = 1.5673$$

$$\text{Mean of the frequency curve} = M_X - S\,M_K = 4.3626 - 1.328\ (1.5673)$$

$$= 2.813$$

IV. Equation of frequency curve (regression line).

$$X = 1.328\ K + 2.813$$

Figure 4.13 Calculation of modified regression line that define the range of events from both curves and the corresponding exceedance probabilities are picked off the frequency curve and used to compute the exceedance probability of the combined curve. In this example an additional step is used to obtain the greatest possible accuracy. The normal (K_p) and Pearson Type III ($K_{G,P}$) deviates are computed using Euqations 4.1 and 4.2 for the hurricane and non-hurricane curves, respectively.

$$K_p = \frac{\log Q - \overline{X}_H}{S_H} \tag{4.1}$$

$$K_{G,P} = \frac{\log Q - \overline{X}_N}{S_N} \tag{4.2}$$

where $\overline{X}_H$ and $\overline{X}_N$ are mean of the logarithms for the hurricane and non-hurricane events respectively, and S_H and S_N are the standard deviations of the logarithms for the hurricane and non-hurricane events respectively. The exceedance probabilities corresponding to these deviates can be found using tables of the normal or Pearson type III distributions (Benjamin and Cornell, 1970; U.S. Water Resources Council, 1981).

Figure 4.14 summarizes the calculations to develop a combined-population frequency curve for West Conewago Creek. Column 1 contains the selected discharges and columns 2 and 4 contain the normal and Pearson type III deviates calculated using Equations 4.1 and 4.2, respectively. The corresponding exceedance probabilities are used in Equation 3.4 to calculate the combined-population exceedance probability in column 6 that corresponds to the discharge tabulated in column 1.

Q (1)	Hurricane K_P (2)	Hurricane P_H (3)	Non-hurricane $K_{G,P}$ (4)	Non-hurricane P_N (5)	Combined P_C (6)	Combined K_P (7)	W_N (8)	W_H (9)	S_C (10)	G_C (11)	Combined $K_{G,P}$ (12)	a (13)	b (14)	$\sqrt{(K_{G,P})^2-ab}$ (15)	$K^U_{P,c}$ (16)	$K^L_{P,c}$ (17)	$E^U_{P,c}$ (18)	$E^L_{P,c}$ (19)	$Q^U_{P,c}$ (20)	$Q^L_{P,c}$ (21)
100000	2.3271	0.010	6.2774	<0.0001	≈0.01	2.3271	0	1.0	0.8710	0	2.3271	0.9685	5.3539	0.4795	2.8978	1.9076	0.5707	-0.4195	314109	43114
50000	1.9814	0.0238	4.0141	<0.0001	≈0.0238	1.9814	0	1.0	0.8710	0	1.9814	0.9685	3.8645	0.4279	2.4875	1.6040	0.5061	-0.3774	137970	23456
30000	1.7267	0.0421	2.3460	<0.0001	≈0.0421	1.7267	0	1.0	0.8710	0	1.7267	0.9685	2.9200	0.3916	2.1871	1.3785	0.4604	-0.3482	75532	14922
26500	1.6649	0.0480	1.9409	0.0022	0.0501	1.6438	0.0438	0.9562	0.8387	-0.0350	1.6338	0.9685	2.6078	0.3789	2.0780	1.2957	0.4442	-0.3381	62488	13794
25000	1.6358	0.0509	1.7507	0.0091	0.0595	1.5589	0.1517	0.8483	0.7600	-0.1214	1.5292	0.9685	2.2770	0.3649	1.9556	1.2022	0.4264	-0.3270	52723	14107
20000	1.5246	0.0637	1.0220	0.1528	0.2069	0.8173	0.7058	0.2942	0.3501	-0.5646	0.8355	0.9685	0.6366	0.2855	1.1574	0.5679	0.3219	-0.2676	25926	16119
17500	1.4580	0.0724	0.5860	0.3077	0.3581	0.3635	0.8095	0.1905	0.2736	-0.6476	0.4479	0.9685	0.1391	0.2566	0.7274	0.1975	0.2795	-0.2504	20869	14946
15000	1.3811	0.0836	0.0826	0.5198	0.5599	-0.1507	0.8615	0.1385	0.2352	-0.6892	-0.0360	0.9685	-0.0602	0.2441	0.2149	-0.2892	0.2509	-0.2532	17183	13078
12500	1.2902	0.0983	-0.5127	0.7272	0.7540	-0.6871	0.8809	0.1191	0.2209	-0.7047	-0.6086	0.9685	0.3089	0.2668	-0.3529	-0.9039	0.2557	-0.2953	14236	10757
10000	1.1790	0.1192	-1.2414	0.8829	0.8969	-1.2640	0.8810	0.1190	0.2208	-0.7048	-1.3079	0.9685	1.6491	0.3367	-1.0027	-1.6980	0.3052	-0.3901	11679	8201

Figure 4.14 Calculations for determining the combined-population frequency curve, and confidence limits using Method 1

$\overline{X}_H = 2.9731$ $\overline{X}_N = 4.1651$

$S_H = .87$ $S_N = .1330$

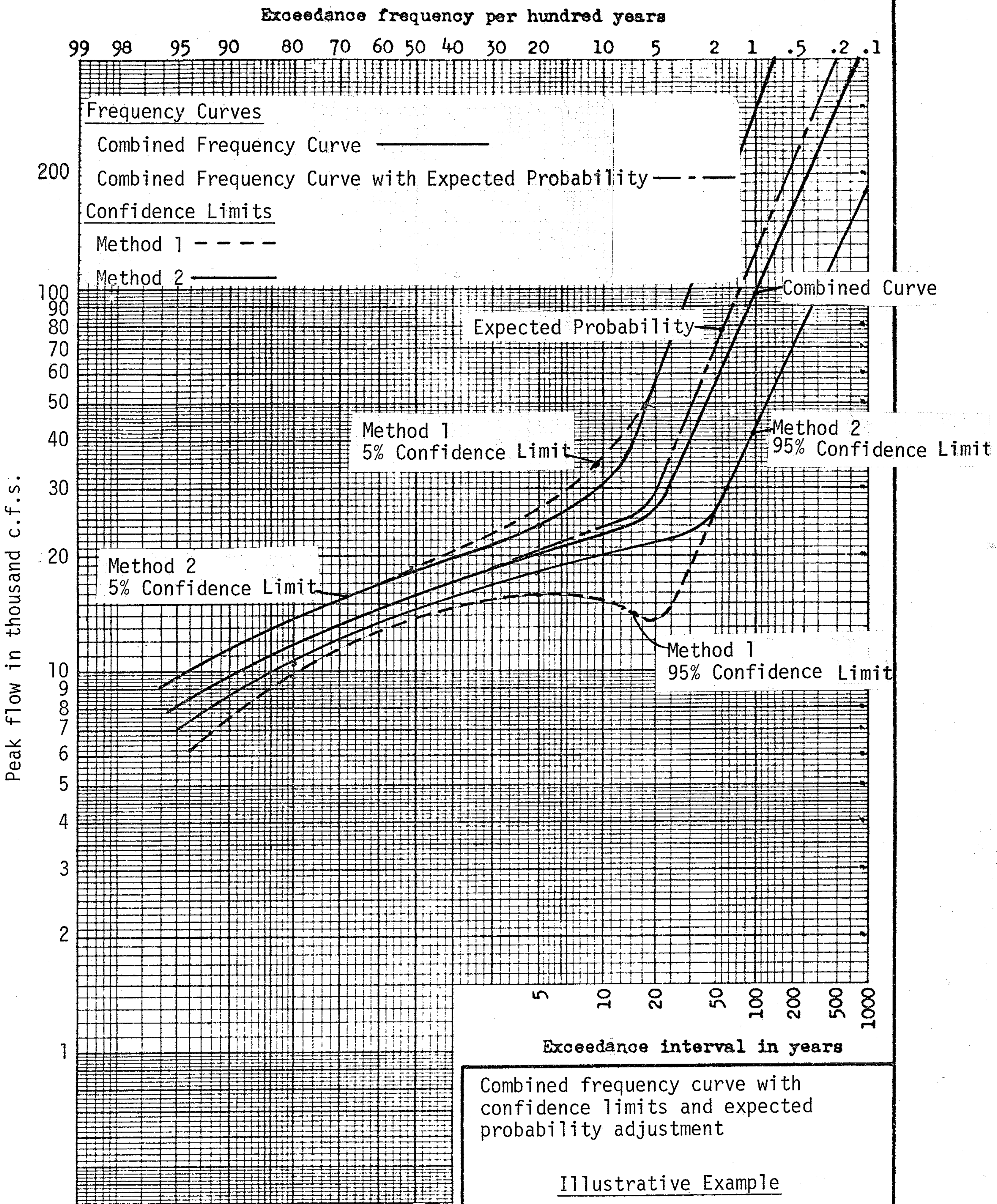
$\overline{G}_N = -.8000$

Section 4.3.4. Expected Probability Calculations

The expected probability adjustment is calculated for the combined-population frequency curve using 44 years of record. An expanded version of Exhibit 40 in "Statistical Methods in Hydrology" (U.S. Army Corps of Engineers, 1962) was used to determine the P_N values as shown below in Figure 4.15, and plotted on Figure 4.16.

P_∞	P_N	Q
0.0100	0.0132	99900
0.0400	0.0453	57800
0.0500	0.0556	26600
0.1000	0.1059	22700
0.3000	0.3034	18300
0.4000	0.4017	16900
0.5000	0.5000	15700
0.6000	0.5983	14500
0.7000	0.6966	13300
0.9000	0.8941	9950
0.9500	0.9444	8590

Figure 4.15 Tabulation of P_N vs P_∞ for the combined-population curve.



Combined frequency curve with confidence limits and expected probability adjustment

Illustrative Example

Figure 4.16

Section 4.3.5. Confidence Limits Calculations

The two methods for calculating confidence limits that were described in Section 3.7 are used in this section to develop 5 and 95 percent confidence limits for West Conewago Creek.

Method 1

Equations 3.5, 3.8, 3.9 and 3.12 are used to determine the confidence limits as illustrated in columns 7 through 21 in Figure 4.14. The first step is to determine equivalent values of the Pearson type III deviates, $K_{G,P}$ which are needed in Equations 9-4 through 9-6 in Bulletin 17B to calculate values of $K_{P,c}$ in Equations 3.8 and 3.9. z_c is the normal deviate that is associated with the level of significance c.

Values of K_p, in column 7 are the equivalent normal deviates corresponding to the probabilities, P_C, in column 6 of the combined curve. In order to get the equivalent Pearson deviate of the combined curve $K_{G,P}$ the equivalent skew coefficient (G_C) must be determined for each discharge. This is accomplished using an equation identical to Equation 3.12, except the skew coefficient of the hurricane and non-hurricane frequency curves are used in lieu of the standard deviations. Column 8 is the relative weight of the non-hurricane probabilities, and column 9 is the relative weight of the hurricane probabilities. These are used to determine the equivalent standard deviation, S_C, in column 10 and the equivalent skew coefficient, G_C, in column 11. The standard deviations and skew coefficients for the hurricane and non-hurricane frequency curves are: 0.1330, -0.8, 0.87 and 0, respectively. The Wilson-Hilferty (1931) approximation shown in Equation 4.3, is used with the equivalent skew coefficient of the combined curve G_C in column 11, and the normal deviate, K_P in column 7 to obtain the equivalent Pearson deviate $K_{G,P}$ shown in column 12.

$$K_{G,P} = \{[K_P - G_C/6)(G_C/6) + 1]^3 - 1\} (2/G_C) \tag{4.3}$$

This Pearson deviate is then used to develop $K^U_{P,c}$ and $K^L_{P,c}$ in columns 16 and 17 using equations 9-4 through 9-6 in Bulletin 17B as shown below for the upper and lower confidence limits respectively.

$$\text{Col. 13} = a = 1 - \frac{z_c}{2(N-1)} \tag{4.4}$$

$$\text{Col. 14} = b = (K_{G,P})^2 - \frac{z_c^2}{N} \tag{4.5}$$

$$\text{Col. 15} = \sqrt{(K_{G,P})^2 - ab} \tag{4.6}$$

$$\text{Col. 16} = K^U_{P,c} = \frac{K_{G,P} + \sqrt{(K_{G,P})^2 - ab}}{a} \tag{4.7}$$

$$\text{Col. 17} = K^L_{P,c} = \frac{K_{G,P} - \sqrt{(K_{G,P})^2 - ab}}{a}$$

The terms $E^L_{P,c}$ and $E^U_{P,c}$ in columns 18 and 19 are found using equations 3.8 and 3.9. These are then used with the standard deviation in column 10 to determine the upper and lower confidence limits $Q^U_{P,C}$ and $Q^L_{P,c}$ with equation 3.5 as shown in columns 20 and 21. These curves are drawn as a dashed line in Figure 4.16, revealing the irregular shape of the lower confidence. This is caused by the large change in standard deviation near the intersection of the hurricane and non-hurricane frequency curves.

Method 2

The second method for developing confidence limits for the combined-population frequency curve requires that separate confidence limits be calculated for the hurricane and non-hurricane frequency curves. Figure 4.17 illustrates these computations. The 5 and 95 percent confidence limits, shown in columns 3, 4, 6, and 7, are calculated using Equations 9-4 through 9-6 in Bulletin 17B (U.S. Water Resources Council, 1981) for the discharges contained in column 1. These confidence limits are drawn on Figure 4.18 and the probabilities associated with the 5 and 95 percent confidence limits (columns 9, 10, 12, and 13) are selected that correspond to the discharge shown in column 8. Finally Equation 3.4 is used to combine the hurricane and non-hurricane curves yielding the combined confidence limit curves which are tabulated in columns 12 and 14 and shown in Figure 4.18.

For this station the confidence limits calculated using Method 2 seem more reasonable than the one calculated using Method 1.

	Hurricane			Non-Hurricane				95% Confidence Limit			5% Confidence Limit		
Q	P_H	$Q^U_{P,c}$	$Q^L_{P,c}$	P_N	$Q^U_{P,c}$	$Q^L_{P,c}$	Q	P_H	P_N	P_C	P_H	P_N	P_C
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
100000	0.01	314089	43115	--	--	--	100000	0.0029	--	≈0.0029	0.032	--	≈0.032
50000	0.0238	137952	23452	--	--	--	50000	0.008	--	≈0.008	0.060	--	≈0.060
30000	0.0421	75521	14919	--	--	--	30000	0.017	--	≈0.017	0.095	0.004	0.0986
26500	0.0480	65289	13360	0.0022	30872	23642	28000	0.019	--	≈0.019	0.100	0.013	0.1117
25000	0.0509	60968	12683	0.0091	28823	22453	26000	0.021	0.0001	0.0211	0.105	0.046	0.1462
20000	0.0637	46960	10389	0.1528	22243	18353	24000	0.023	0.0018	0.0248	0.111	0.080	0.1821
17500	0.0724	40182	9215	0.3077	19142	16186	22000	0.026	0.013	0.0387	0.120	0.170	0.2696
15000	0.0836	33579	8020	0.5198	16219	13895	20000	0.029	0.062	0.0892	0.128	0.280	0.3722
12500	0.0983	27178	6800	0.7272	13504	11453	18000	0.033	0.180	0.2071	0.138	0.410	0.4914
10000	0.1192	21008	5551	0.8829	10959	8902	16000	0.038	0.340	0.3651	0.150	0.550	0.6175
							14000	0.045	0.520	0.5416	0.162	0.700	0.7486
							12000	0.053	0.680	0.6970	0.180	0.825	0.8565
							10000	0.065	0.825	0.8364	0.200	0.925	0.9400

Figure 4.17 Calculations for confidence limits using Method 2

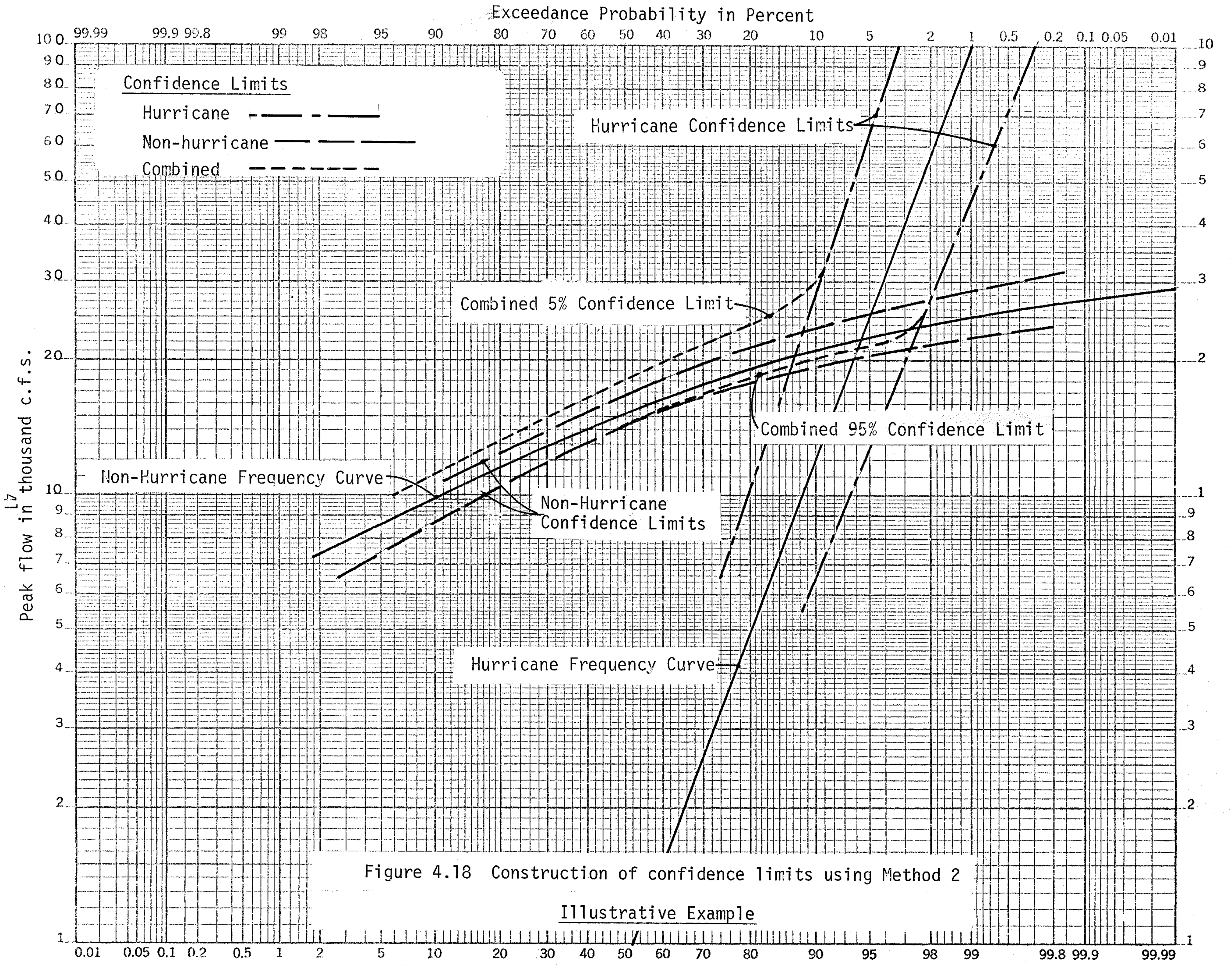


Figure 4.18 Construction of confidence limits using Method 2

Illustrative Example

REFERENCES

Benjamin, J. R., and C. A. Cornell (1970). Probability, Statistics, and Decision for Civil Engineers, McGraw-Hill.

Hardison, C. W. and M. E. Jennings (1972). "Bias in Computed Flood Risk," Journal of the Hydraulics Division - ASCE.

Lloyd, E. (1978). "Expected Probability, What Protection is Actually Achieved by Levees Designed to Withstand a 50-year Flood," Unpublished.

Resnikoff, G. J. and G. J. Lieberman (1957). Tables of the Non-Central t Distribution, Stanford University Press.

U.S. Army Corps of Engineers, Sacramento District (1958). Frequency of New England Floods, Research Note No. 1, Sacramento, California.

U.S. Army Corps of Engineers, Sacramento District (1959). Frequency Curves from Incomplete Samples, Research Note No. 3, Sacramento, California.

U.S. Army Corps of Engineers, Sacramento District (1962). Statistical Methods in Hydrology, Sacramento, California.

U.S. Army Corps of Engineers, Hydrologic Engineering Center (1965). Generalized Frequency Studies for Kanawha River, Huntington District, Special Project Memo No. 45, Davis, California.

U.S. Army Corps of Engineers, Hydrologic Engineering Center (1975a). "Hydrologic Frequency Analysis," IHD Volume 3, Davis, California.

U.S. Army Corps of Engineers, North Atlantic Division (1975b). Hydrologic Study, Tropical Storm Agnes, New York, New York.

U.S. Army Corps of Engineers, Hydrologic Engineering Center (1976). Flood Flow Frequency Analysis, Users Manual, Davis, California.

U.S. Army Corps of Engineers (1980). Hydrologic Frequency Analysis, Draft Engineering Manual, Chapter 10.

U.S. Department of Commerce, Weather Bureau (1965). Tropical Cyclones of the North Atlantic Ocean, Technical Paper No. 55.

U.S. Geological Survey (1968). "Magnitude and Frequency of Floods in the United States, Part 1-B, North Atlantic Slope Basins, New York to York River," Water Supply Paper No. 1672, Reston, Virginia.

U.S. Geological Survey (1978). "Composite Log-Type III Frequency-Magnitude Curve of Annual Floods," by Crippen, John R., Open-File Report 78-352, Menlo Park, California.

U.S. Water Resources Council (1981). Guidelines for Determining Flood Flow Frequency, Bulletin 17B.

Wilson, E. B., and M. M. Hilferty (1931). "The Distribution of Chi-Square," Proc., National Academy of Science, Vol. 17, No. 12, pp 684-688.

Appendix I

DERIVATION OF EQUATIONS 3.3 & 3.4

- Fundamental equation representing the combined probability of two independent probabilities is:

$$P_c = P_1 \times P_2$$

- This represents, however, the probability that both P_1 and P_2 will occur. For the application of combining two frequency curves, the question is whether one or the other event will occur.

- The probability of occurrence of one or another event is equivalent to [1 - (Probability of both not occurring)]

- Probability of "non-occurrence" for P_1 is $1 - P_1$ and for P_2 is $1 - P_2$

- Combined probability of both not occurring = $(1 - P_1)(1 - P_2)$

- Probability of either/or occurring = $1 - (1 - P_1)(1 - P_2)$

  This is the form of equation 3.3.

- Equation 3.4 is found by algebra:

$$P_c = 1 - (1 - P_1)(1 - P_2)$$

$$= 1 - (1 - P_1 - P_2 + P_1 P_2)$$

$$= P_1 + P_2 - P_1 P_2$$