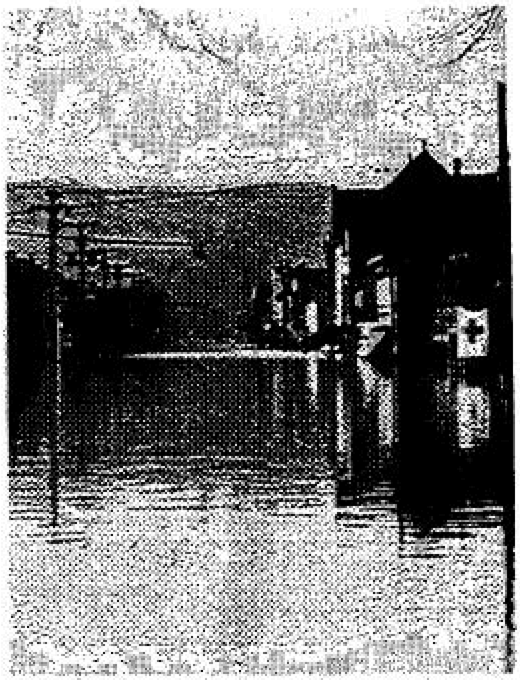

Guidelines For Determining

Flood Flow Frequency

Bulletin # 17B
of the
Hydrology Subcommittee

Revised September 1981
Editorial Corrections March 1982

INTERAGENCY ADVISORY COMMITTEE
ON WATER DATA

U.S. Department of the Interior
Geological Survey
Office of Water Data Coordination
Reston, Virginia 22092

FOREWORD

An accurate estimate of the flood damage potential is a key element to an effective, nationwide flood damage abatement program. Further, there is an acute need for a consistent approach to such estimates because management of the nation's water and related land resources is shared among various levels of government and private enterprise. To obtain both a consistent and accurate estimate of flood losses requires development, acceptance, and widespread application of a uniform, consistent and accurate technique for determining flood-flow frequencies.

In a pioneer attempt to promote a consistent approach to flood-flow frequency determination, the U.S. Water Resources Council in December 1967 published Bulletin No. 15, "A Uniform Technique for Determining Flood Flow Frequencies." The technique presented therein was adopted by the Council for use in all Federal planning involving water and related land resources. The Council also recommended use of the technique by State, local government, and private organizations. Adoption was based upon the clear understanding that efforts to develop methodological improvements in the technique would be continued and adopted when appropriate.

An extension and update of Bulletin No. 15 was published in March 1976 as Bulletin No. 17, "Guidelines for Determining Flood Flow Frequency." It presented the currently accepted methods for analyzing peak flow frequency data at gaging stations with sufficient detail to promote uniform application. The guide was a synthesis of studies undertaken to find methodological improvements and a survey of existing literature on peak flood flow determinations.

* The present guide is the second revision of the original publication. *

and improves the methodologies. It revises and expands some of the techniques in the previous editions of this Bulletin and offers a further explanation of other techniques. It is the result of a continuing effort to develop a coherent set of procedures for accurately defining flood potentials. Much additional study is required before the two goals of accuracy and consistency will be fully attained. All who are interested in improving peak flood-flow frequency determinations are encouraged to submit comments, criticism and proposals to the Office of Water Data Coordination for consideration by the Hydrology Subcommittee.

Federal agencies are requested to use these guidelines in all planning activities involving water and related land resources. State, local and private organizations are encouraged to use these guidelines also to assure more uniformity, compatibility, and comparability in the frequency values that all concerned agencies and citizens must use for many vital decisions.

This present revision is adopted with the knowledge and understanding that review of these procedures will continue. When warranted by experience and by examination and testing of new techniques, other revisions will be published.

HYDROLOGY SUBCOMMITTEE [1/]

Member	Agency	Department
* Robert E. Rallison	Soil Conservation Service	Agriculture
Robert G. Delk	Forest Service	"
Walter J. Rawls	Science Education Administration	"
Vernon K. Hagen	Corps of Engineers	Army
Roy G. Huffman	"	"
Allen F. Flanders	NOAA, National Weather Service	Commerce
John F. Miller	"	"
Truman Goins		Housing and Urba Development
Porter Ward	Geological Survey	Interior
David F. Gudgel	Bureau of Reclamation	"
Don Willen	Office of Surface Mining	"
Ewell H. Mohler, Jr.	Office of Water Research and Technology	"
Sidney J. Spiegel	Bureau of Indian Affairs	"
Pat Diffy	Bureau of Mines	"
Leo Fake	Fish and Wildlife Service	"
Victor Berte'	National Park Service	"
Irene L. Murphy	Heritage, Conservation and Recreation Service	"
D. C. Woo	Federal Highway Administration	Transportation
Philip L. Thompson	"	"
Timothy Stuart		Environmental Protection Agenc
Robert Horn		"
Steve Parker		Federal Energy Regulatory Commission
Patrick Jefferson		"
Brian Mrazik		Federal Emergenc Management Agenc
William S. Bivins		"
Edward F. Hawkins		Nuclear Regulatory Commission

HYDROLOGY SUBCOMMITTEE [1/] - Con't

Member	Agency	Department
*Donald W. Newton		Tennessee Valley Authority
Larry M. Richardson		"
Ron Scullin		Water Resources Council

WORK GROUP ON REVISION OF BULLETIN 17

Member	Agency	Department
Roger Cronshey	Soil Conservation Service	Agriculture
Roy G. Huffman	Corps of Engineers	Army
John F. Miller*	NOAA, National Weather Service	Commerce
William H. Kirby	Geological Survey	Interior
Wilbert O. Thomas, Jr.	"	"
Frederick A. Bertle	Bureau of Reclamation	"
Donald W. Newton		Tennessee Valley Authority *

* Chairman

[1/] Membership as of September 1981

The following pages contain revisions from material presented in Bulletin 17, "Guidelines for Determining Flood Flow Frequency."

1, 4, 8-2, and 13-1

The revised material is included on the lines enclosed by the + symbol.

The following pages of Bulletin 17 have been deleted:

13-2 through 13-35

The following pages contain revisions from the material in either Bulletin 17 or 17A.

i, ii, iii, iv, v, vi, vii, 1, 3, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 26, 1-1, 1-2, 1-3, 1-4, 2-3, 2-7, 2-8, 4-1, 5-1, 5-2, 5-3, 5-4, 6-1, 6-2, 6-3, 6-5, 6-6, 6-7, 7-1, 7-2, 7-3, 7-4, 7-5, 7-6, 7-7, 7-8, 7-9, 9-1 through 9-10, 10-1, 10-2, 10-3, 12-2 through 12-37 and 14-1

The revised material is included on the lines enclosed by the * symbol.

The following page of Bulletin 17 and 17A has been deleted from 17B: 4-2

Editorial corrections to Bulletin 17B were incorporated into this report in March 1982.

CONTENTS

*

I. Introduction

In December 1967, Bulletin No. 15, "A Uniform Technique for Determining Flood Flow Frequencies," was issued by the Hydrology Committee of the Water Resources Council. The report recommended use of the Pearson Type III distribution with log transformation of the data (log-Pearson Type III distribution) as a base method for flood flow frequency studies. As pointed out in that report, further studies were needed covering various aspects of flow frequency determinations.

In March 1976, Bulletin 17, "Guidelines for Determining Flood Flow Frequency" was issued by the Water Resources Council. The guide was an extension and update of Bulletin No. 15. It provided a more complete guide for flood flow frequency analysis incorporating currently accepted technical methods with sufficient detail to promote uniform application. It was limited to defining flood potentials in terms of peak discharge and exceedance probability at locations where a systematic record of peak flood flows is available. The recommended set of procedures was selected from those used or described in the literature prior to 1976, based on studies conducted for this purpose at the Center for Research in Water Resources of the University of Texas at Austin (summarized in Appendix 14) and on studies by the Work Group on Flood Flow Frequency.

The "Guidelines" were revised and reissued in June 1977 as Bulletin 17A. Bulletin 17B is the latest effort to improve and expand upon the earlier publications. Bulletin 17B provides revised procedures for weighting a station skew value with the results from a generalized skew study, detecting and treating outliers, making two station comparisons, and computing confidence limits about a frequency curve. The Work Group that prepared this revision did not address the suitability of the original distribution or the generalized skew map.

Major problems are encountered when developing guides for flood flow frequency determinations. There is no procedure or set of procedures that can be adopted which, when rigidly applied to the available data, will accurately define the flood potential of any given watershed. Statistical analysis alone will not resolve all flood frequency problems. As discussed in subsequent sections of this guide, elements of risk and uncertainty are inherent in any flood frequency analysis. User decisions must be based on properly applied procedures and proper interpretation of results considering risk and uncertainty. Therefore, the judgment of a professional experienced in hydrologic analysis will enhance the usefulness of a flood frequency analysis and promote appropriate application.

It is possible to standarize many elements of flood frequency analysis. This guide describes each major element of the process of defining the flood potential at a specific location in terms of peak discharge and exceedance probability. Use is confined to stations where available records are adequate to warrant statistical analysis of the data. Special situations may require other approaches. In those cases where the procedures of this guide are not followed, deviations must be supported by appropriate study and accompanied by a comparison of results using the recommended procedures.

As a further means of achieving consistency and improving results, the Work Group recommends that studies be coordinated when more than one analyst is working currently on data for the same location. This recommendation holds particularly when defining exceedance probabilities for rare events, where this guide allows more latitude.

Flood records are limited. As more years of record become available at each location, the determination of flood potential may change. Thus, an estimate may be outdated a few years after it is made. Additional flood data alone may be sufficient reason for a fresh assessment of the flood potential. When making a new assessment, the analyst should incorporate in his study a review of earlier estimates. Where differences appear, they should be acknowledged and explained.

II. Summary

This guide describes the data and procedures for computing flood flow frequency curves where systematic stream gaging records of sufficient length (at least 10 years) to warrant statistical analysis are available as the basis for determination. The procedures do not cover watersheds where flood flows are appreciably altered by reservoir regulation or where the possibility of unusual events, such as dam failures, must be considered. The guide was specifically developed for the treatment of annual flood peak discharge. It is recognized that the same techniques could also be used to treat other hydrologic elements, such as flood volumes. Such applications, however, were not evaluated and are not intended.

The guide is divided into six broad sections which are summarized below:

A. Information to be Evaluated

The following categories of flood data are recognized: systematic records, historic data, comparison with similar watersheds, and flood estimates from precipitation. How each can be used to define the flood potential is briefly described.

B. Data Assumptions

A brief discussion of basic data assumptions is presented as a reminder to those developing flood flow frequency curves to be aware of potential data errors. Natural trends, randomness of events, watershed changes, mixed populations, and reliability of flow estimates are briefly discussed.

C. Determination of the Frequency Curve

This section provides the basic guide for determination of the frequency curve. The main thrust is determination of the annual flood series. Procedures are also recommended to convert an annual to partial-duration flood series.

The Pearson Type III distribution with log transformation of the flood data (log-Pearson Type III) is recommended as the basic distribution for defining the annual flood series. The method of moments is used to determine the statistical parameters of the distribution from station data.
*Generalized relations are used to modify the station skew coefficient. *
Methods are proposed for treatment of most flood record problems encoun-
*tered. Procedures are described for refining the basic curve determined from statistical analysis of the systematic record and historic flood data to incorporate information gained from comparisons with similar watersheds *
and flood estimates from precipitation.

D. Reliability Applications

Procedures for computing confidence limits to the frequency curve are provided along with those for calculating risk and for making expected probability adjustments.

E. Potpourri

This section provides information of interest but not essential to the guide, including a discussion of non-conforming special situations, plotting positions, and suggested future studies.

F. Appendix

The appendix provides a list of references, a glossary and list of symbols, tables of K values, the computational details for treating most of the recommended procedures, information about how to obtain a computer program for handling the statistical analysis and treatment of data, and a summary of the report ("Flood Flow Frequency Techniques") describing studies made at the University of Texas which guided selection of some of the procedures proposed.

III. Information to be Evaluated

When developing a flood flow frequency curve, the analyst should consider all available information. The four general types of data which can be included in the flood flow frequency analysis are described in the following paragraphs. Specific applications are discussed in subsequent sections.

A. Systematic Records

Annual peak discharge information is observed systematically by many Federal and state agencies and private enterprises. Most annual peak records are obtained either from a continuous trace of river stages or from periodic observations of a crest-stage gage. Crest-stage records may provide information only on peaks above some preselected base. A major portion of these data are available in U.S. Geological Survey (USGS) Water Supply Papers and computer files, but additional information in published or unpublished form is available from other sources.

A statistical analysis of these data is the primary basis for the determination of the flow frequency curve for each station.

B. Historic Data

At many locations, particularly where man has occupied the flood plain for an extended period, there is information about major floods which occurred either before or after the period of systematic data collection. This information can often be used to make estimates of peak discharge. It also often defines an extended period during which the largest floods, either recorded or historic, are known. The USGS includes some historic flood information in its published reports and computer files. Additional information can sometimes be obtained from the files of other agencies or extracted from newspaper files or by intensive inquiry and investigation near the site for which the flood frequency information is needed.

Historic flood information should be obtained and documented whenever possible, particularly where the systematic record is relatively short. Use of historic data assures that estimates fit community experience and improves the frequency determinations.

C. Comparison With Similar Watersheds

Comparisons between computed frequency curves and maximum flood data of the watershed being investigated and those in a hydrologically similar region are useful for identification of unusual events and for testing the reasonableness of flood flow frequency determinations. Studies have been made and published [e.g., (1), (2), (3), (4)]* which permit comparing flood frequency estimates at a site with generalized estimates for a homogeneous region. Comparisons with information at stations in the immediate region should be made, particularly at gaging stations upstream and downstream, to promote regional consistency and help prevent gross errors.

*Numbers in parentheses refer to numbered references in Appendix 1.

D. Flood Estimates From Precipitation

Flood discharges estimated from climatic data (rainfall and/or snowmelt) can be a useful adjunct to direct streamflow measurements. Such estimates, however, require at least adequate climatic data and a valid watershed model for converting precipitation to discharge. Unless such models are already calibrated to the watershed, considerable effort may be required to prepare such estimates.

Whether or not such studies are made will depend upon the availabilit of the information, the adequacy of the existing records, and the exceedar probability which is most important.

IV. Data Assumptions

Necessary assumptions for a statistical analysis are that the array of flood information is a reliable and representative time sample of random homogeneous events. Assessment of the adequacy and applicability of flood records is therefore a necessary first step in flood frequency analysis. This section discusses the effect of climatic trends, randomnes of events, watershed changes, mixed populations, and reliability of flow estimates on flood frequency analysis.

A. Climatic Trends

There is much speculation about climatic changes. Available evidence indicates that major changes occur in time scales involving thousands of years. In hydrologic analysis it is conventional to assume flood flows are not affected by climatic trends or cycles. Climatic time invariance was assumed when developing this guide.

B. Randomness of Events

In general, an array of annual maximum peak flow rates may be considered a sample of random and independent events. Even when statistical tests of the serial correlation coefficients indicate a significant deviation from this assumption, the annual peak data may define an unbiase estimation of future flood activity if other assumptions are attained. The nonrandomness of the peak series will, however, increase the degree of uncertainty in the relation; that is, a relation based upon nonrandom data will have a degree of reliability attainable from a lesser sample of random data (5), (6).

C. Watershed Changes

It is becoming increasingly difficult to find watersheds in which the flow regime has not been altered by man's activity. Man's activities which can change flow conditions include urbanization, channelization, levees, the construction of reservoirs, diversions, and alteration of cover conditions.

Watershed history and flood records should be carefully examined to assure that no major watershed changes have occurred during the period of record. Documents which accompany flood records often list such changes. All watershed changes which affect record homogeneity, however, might not be listed; unlisted, for instance, might be the effects of urbanization and the construction of numerous small reservoirs over a period of several years. Such incremental changes may not significantly alter the flow regime from year to year but the cumulative effect can after several years.

Special effort should be made to identify those records which are not homogeneous. Only records which represent relatively constant watershed conditions should be used for frequency analysis.

D. Mixed Populations

At some locations flooding is created by different types of events. For example, flooding in some watersheds is created by snowmelt, rainstorms, or by combinations of both snowmelt and rainstorms. Such a record may not be homogeneous and may require special treatment.

E. Reliability of Flow Estimates

Errors exist in streamflow records, as in all other measured values. Errors in flow estimates are generally greatest during maximum flood flows. Measurement errors are usually random, and the variance introduced is usually small in comparison to the year-to-year variance in flood flows. The effects of measurement errors, therefore, may normally be neglected in flood flow frequency analysis. Peak flow estimates of historic floods can be substantially in error because of the uncertainty in both stage and stage-discharge relationships.

At times errors will be apparent or suspected. If substantial, the errors should be brought to the attention of the data collecting agency with supporting evidence and a request for a corrected value. A more complete discussion of sources of error in streamflow measurement is found in (7).

V. Determination of Frequency Curve

A. Series Selection

Flood events can be analyzed using either annual or partial-duration series. The annual flood series is based on the maximum flood peak for each year. A partial-duration series is obtained by taking all flood peaks equal to or greater than a predefined base flood.

If more than one flood per year must be considered, a partial-duration series may be appropriate. The base is selected to assure that all events of interest are evaluated including at least one event per time period. A major problem encountered when using a partial-duration series is to define flood events to ensure that all events are independent. It is common practice to establish an empirical basis for separating flood events. The basis for separation will depend upon the investigator and the intended use. No specific guidelines are recommended for defining flood events to be included in a partial series.

A study (8) was made to determine if a consistent relationship existed between the annual and partial series which could be used to convert from the annual to the partial-duration series. Based on this study as summarized in Appendix 14, the Work Group recommends that the partial-duration series be developed from observed data. An alternative but less desirable solution is to convert from the annual to the partial-duration series. For this, the first choice is to use a conversion factor specifically developed for the hydrologic region in which the gage is located. The second choice is to use published relationships [e.g., (9)].

Except for the preceding discussion of the the partial-duration series, the procedures described in this guide apply to the annual flood series.

B. Statistical Treatment

1. The Distribution--Flood events are a succession of natural events which, as far as can be determined, do not fit any one specific known statistical distribution. To make the problem of defining flood probabilities tractable it is necessary, however, to assign a distribution. Therefore, a study was sponsored to find which of many possible distributions and alternative fitting methods would best meet the purposes of this guide. This study is summarized in Appendix 14. The Work Group concluded from this and other studies that the Pearson Type III distribution with log transformation of the data (log-Pearson Type III distribution) should be the base method for analysis of annual series data using a generalized skew coefficient as described in the following section.

2. Fitting the Distribution--The recommended technique for fitting a log-Pearson Type III distribution to observed annual peaks is to compute the base 10 logarithms of the discharge, Q, at selected exceedance probability, P, by the equation:

$$\text{Log } Q=\overline{X}+KS \qquad (1)$$

where $\overline{X}$ and S are as defined below and K is a factor that is a function of the skew coefficient and selected exceedance probability. Values of K can be obtained from Appendix 3.

The mean, standard deviation and skew coefficient of station data may be computed using the following equations:

$$\overline{X} = \frac{\Sigma X}{N} \qquad (2)$$

$$S = \left[\frac{\Sigma(X - \overline{X})^2}{(N - 1)}\right]^{0.5} \qquad (3a)$$

$$= \left[\frac{(\Sigma X^2) - (\Sigma X)^2/N}{(N - 1)}\right]^{0.5} \qquad (3b)$$

$$G = \frac{N \Sigma (X - \overline{X})^3}{(N - 1)(N - 2)S^3} \qquad (4a)$$

$$= \frac{N^2(\Sigma X^3) - 3N(\Sigma X)(\Sigma X^2) + 2(\Sigma X)^3}{N(N-1)(N-2)S^3} \qquad (4b)$$

in which:

X = logarithm of annual peak flow
N = number of items in data set
$\overline{X}$ = mean logarithm
S = standard deviation of logarithms
G = skew coefficient of logarithms

Formulas for computing the standard errors for the statistics $\overline{X}$, S, and G are given in Appendix 2. The precision of values computed with equations 3b and 4b is more sensitive than with equations 3a and 4a to the number of significant digits used in their calculation. When the available computation facilities only provide for a limited number of significant digits, equations 3a and 4a are preferable.

* 3. Estimating Generalized Skew--The skew coefficient of the station record (station skew) is sensitive to extreme events; thus it is difficult to obtain accurate skew estimates from small samples. The accuracy of the estimated skew coefficient can be improved by weighting the station skew with generalized skew estimated by pooling information from nearby sites. The following guidelines are recommended for estimating generalized skew.*

*Guidelines on weighting station and generalized skew are provided in the next section of this bulletin.

The recommended procedure for developing generalized skew coefficients requires the use of at least 40 stations, or all stations within a 100-mile radius. The stations used should have 25 or more years of record. It is recognized that in some locations a relaxation of these criteria may be necessary. The actual procedure includes analysis by three methods: 1) skew isolines drawn on a map; 2) skew prediction equation; and 3) the mean of the station skew values. Each of the methods are discussed separately.

To develop the isoline map, plot each station skew value at the centroid of its drainage basin and examine the plotted data for any geographic or topographic trends. If a pattern is evident, then isolines are drawn and the average of the squared differences between observed and isoline values, mean-square error (MSE), is computed. The MSE will be used in appraising the accuracy of the isoline map. If no pattern is evident, then an isoline map cannot be drawn and is therefore, not further considered.

A prediction equation should be developed that would relate either the station skew coefficients or the differences from the isoline map to predictor variables that affect the skew coefficient of the station record. These would include watershed and climatologic variables. The prediction equation should preferably be used for estimating the skew coefficient at stations with variables that are within the range of data used to calibrate the equation. The MSE (standard error of estimate squared) will be used to evaluate the accuracy of the preciction equation.

Determine the arithmetic mean and variance of the skew coefficients for all stations. In some cases the variability of the runoff regime may be so large as to preclude obtaining 40 stations with reasonably homogeneous hydrology. In these situations, the arithmetic mean and variance of about 20 stations may be used to estimate the generalized skew coefficient. The drainage areas and meteorologic, topographic, and geologic characteristics should be representative of the region around the station of interest.

Select the method that provides the most accurate skew coefficient *

* estimates. Compare the MSE from the isoline map to the MSE for the prediction equation. The smaller MSE should then be compared to the variance of the data. If the MSE is significantly smaller than the variance, the method with the smaller MSE should be used and that MSE used in equation 5 for $MSE_{\overline{G}}$. If the smaller MSE is not significantly smaller than the variance, neither the isoline map nor the prediction equation provides a more accurate estimate of the skew coefficient than does the mean value. The mean skew coefficient should be used as it provides the most accurate estimate and the variance should be used in equation 5 for $MSE_{\overline{G}}$.

In the absence of detailed studies the generalized skew ($\overline{G}$) can be read from Plate I found in the flyleaf pocket of this guide. This map of generalized skew was developed when this bulletin was first introduced and has not been changed. The procedures used to develop the statistical analysis for the individual stations do not conform in all aspects to the procedures recommended in the current guide. However, Plate I is still considered an alternative for use with the guide for those who prefer not to develop their own generalized skew procedures.

The accuracy of a regional generalized skew relationship is generally not comparable to Plate I accuracy. While the average accuracy of Plate I is given, the accuracy of subregions within the United States are not given. A comparison should only be made between relationships that cover approximately the same geographical area. Plate I accuracy would be directly comparable to other generalized skew relationships that are applicable to the entire country.

4. <u>Weighting the Skew Coefficient</u>--The station and generalized skew coefficient can be combined to form a better estimate of skew for a given watershed. Under the assumption that the generalized skew is unbiased and independent of station skew, the mean-square error (MSE) of the weighted estimate is minimized by weighting the station and generalized skew in inverse proportion to their individual mean-square errors. This concept is expressed in the following equation adopted from Tasker (39) which should be used in computing a weighted skew coefficient:

$$G_W = \frac{MSE_{\overline{G}}(G) + MSE_G(\overline{G})}{MSE_{\overline{G}} + MSE_G}$$

* where G_W = weighted skew coefficient

G = station skew

$\overline{G}$ = generalized skew

$MSE_{\overline{G}}$ = mean-square error of generalized skew

MSE_G = mean-square error of station skew

Equation 5 can be used to compute a weighted skew estimate regardless of the source of generalized skew, provided the MSE of the generalized skew can be estimated. When generalized skews are read from Plate I, the value of $MSE_{\overline{G}} = 0.302$ should be used in equation 5. The MSE of the station skew for log-Pearson Type III random variables can be obtained from the results of Monte Carlo experiments by Wallis, Matalas, and Slack (40). Their results show that the MSE of the logarithmic station skew is a function of record length and population skew. For use in calculating G_W, this function (MSE_G) can be approximated with sufficient accuracy by the equation:

$$MSE_G \doteq 10^{\left[A - B\left[\log_{10}(N/10)\right]\right]} \qquad (6)$$

Where $A = -0.33 + 0.08|G|$ if $|G| \leq 0.90$
$\qquad\;\; -0.52 + 0.30|G|$ if $|G| > 0.90$

$B = 0.94 - 0.26|G|$ if $|G| \leq 1.50$
$\qquad 0.55$ if $|G| > 1.50$

in which $|G|$ is the absolute value of the station skew (used as an estimate of population skew) and N is the record length in years. If the historic adjustment described in Appendix 6 has been applied, the historically adjusted skew, $\tilde{G}$, and historic period, H, are to be used for G and N, respectively, in equation 6. For convenience in manual computations, equation 6 was used to produce table 1 which shows MSE_G values for selected record lengths and station skews.

*

TABLE 1. - SUMMARY OF MEAN SQUARE ERROR OF STATION SKEW AS A FUNCTION OF RECORD LENGTH AND STATION SKEW. *

STATION SKEW (G OR $\tilde{G}$)	RECORD LENGTH, IN YEARS (N OR H) 10	20	30	40	50	60	70	80	90	100
0.0	0.468	0.244	0.167	0.127	0.103	0.087	0.075	0.066	0.059	0.054
0.1	0.476	0.253	0.175	0.134	0.109	0.093	0.080	0.071	0.064	0.058
0.2	0.485	0.262	0.183	0.142	0.116	0.099	0.086	0.077	0.069	0.063
0.3	0.494	0.272	0.192	0.150	0.123	0.105	0.092	0.082	0.074	0.068
0.4	0.504	0.282	0.201	0.158	0.131	0.113	0.099	0.089	0.080	0.073
0.5	0.513	0.293	0.211	0.167	0.139	0.120	0.106	0.095	0.087	0.079
0.6	0.522	0.303	0.221	0.176	0.148	0.128	0.114	0.102	0.093	0.086
0.7	0.532	0.315	0.231	0.186	0.157	0.137	0.122	0.110	0.101	0.093
0.8	0.542	0.326	0.243	0.196	0.167	0.146	0.130	0.118	0.109	0.100
0.9	0.562	0.345	0.259	0.211	0.181	0.159	0.142	0.130	0.119	0.111
1.0	0.603	0.376	0.285	0.235	0.202	0.178	0.160	0.147	0.135	0.126
1.1	0.646	0.410	0.315	0.261	0.225	0.200	0.181	0.166	0.153	0.143
1.2	0.692	0.448	0.347	0.290	0.252	0.225	0.204	0.187	0.174	0.163
1.3	0.741	0.486	0.383	0.322	0.281	0.252	0.230	0.212	0.197	0.185
1.4	0.794	0.533	0.422	0.357	0.314	0.283	0.259	0.240	0.224	0.211
1.5	0.851	0.581	0.465	0.397	0.351	0.318	0.292	0.271	0.254	0.240
1.6	0.912	0.623	0.498	0.425	0.376	0.340	0.313	0.291	0.272	0.257
1.7	0.977	0.667	0.534	0.456	0.403	0.365	0.335	0.311	0.292	0.275
1.8	1.047	0.715	0.572	0.489	0.432	0.391	0.359	0.334	0.313	0.295
1.9	1.122	0.766	0.613	0.523	0.463	0.419	0.385	0.358	0.335	0.316
2.0	1.202	0.821	0.657	0.561	0.496	0.449	0.412	0.383	0.359	0.339
2.1	1.288	0.880	0.704	0.601	0.532	0.481	0.442	0.410	0.385	0.363
2.2	1.380	0.943	0.754	0.644	0.570	0.515	0.473	0.440	0.412	0.389
2.3	1.479	1.010	0.808	0.690	0.610	0.552	0.507	0.471	0.442	0.417
2.4	1.585	1.083	0.866	0.739	0.654	0.592	0.543	0.505	0.473	0.447
2.5	1.698	1.160	0.928	0.792	0.701	0.634	0.582	0.541	0.507	0.479
2.6	1.820	1.243	0.994	0.849	0.751	0.679	0.624	0.580	0.543	0.513
2.7	1.950	1.332	1.066	0.910	0.805	0.728	0.669	0.621	0.582	0.550
2.8	2.089	1.427	1.142	0.975	0.862	0.780	0.716	0.666	0.624	0.589
2.9	2.239	1.529	1.223	1.044	0.924	0.836	0.768	0.713	0.669	0.631
3.0	2.399	1.638	1.311	1.119	0.990	0.895	0.823	0.764	0.716	0.676

*

* Application of equation 6 and table 1 to stations with absolute skew values (logs) greater than 2 and long periods of record gives relatively little weight to the station value. Application of equation 5 may also give improper weight to the generalized skew if the generalized and station skews differ by more than 0.5. In these situations, an examination of the data and the flood-producing characteristics of the watershed should be made and possibly greater weight given to the station skew. *

5. <u>Broken Record</u>--Annual peaks for certain years may be missing because of conditions not related to flood magnitude, such as gage removal. In this case, the different record segments are analyzed as a continuous record with length equal to the sum of both records, unless there is some physical change in the watershed between segments which may make the total record nonhomogeneous.

6. <u>Incomplete Record</u>--An incomplete record refers to a streamflow record in which some peak flows are missing because they were too low or too high to record, or the gage was out of operation for a short period because of flooding. Missing high and low data require different treatment.

When one or more high annual peaks during the period of systematic record have not been recorded, there is usually information available from which the peak discharge can be estimated. In most instances the data collecting agency routinely provides such estimates. If not, and such an estimate is made as part of the flood frequency analysis, it should be documented and the data collection agency advised.

At some crest gage sites the bottom of the gage is not reached *in some years. For this situation use of the conditional probability adjustment is recommended as described in Appendix 5. *

7. <u>Zero Flood Years</u>--Some streams in arid regions have no flow for the entire year. Thus, the annual flood series for these streams will have one or more zero flood values. This precludes the normal statistical analysis of the data using the recommended log-Pearson Type III *distribution because the logarithm of zero is minus infinity. The conditional probability adjustment is recommended for determining frequency curves for records with zero flood years as described in Appendix 5. *

8. Mixed Population--Flooding in some watersheds is created by different types of events. This results in flood frequency curves with abnormally large skew coefficients reflected by abnormal slope changes when plotted on logarithmic normal probability paper. In some situations the frequency curve of annual events can best be described by computing separate curves for each type of event. The curves are then combined.

Two examples of combinations of different types of flood-producing events include: (1) rain with snowmelt and (2) intense tropical storms with general cyclonic storms. Hydrologic factors and relationships operating during general winter rain flood are usually quite different from those operating during spring snowmelt floods or during local summer cloudburst floods. One example of mixed population is in the Sierra Nevada region of California. Frequency studies there have been 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. Peak flows were segregated by cause--those predominately caused by snowmelt and those predominately caused by rain. Another example is along the Atlantic and Gulf Coasts, where in some instances floods from hurricane and nonhurricane events have been separated, thereby improving frequency estimates.

When it can be shown that there are two or more distinct and genera independent causes of floods it may be more reliable to segregate the flood data by cause, analyze each set separately, and then to combine the data sets using procedures such as described in (11). Separation by calendar periods in lieu of separation by events is not considered hydrologically reasonable unless the events in the separate periods are clearly caused by different hydrometeorologic conditions. The fitting procedures of this guide can be used to fit each flood series separately with the exception that generalized skew coefficients cannot be used unless developed for specific type events being examined.

If the flood events that are believed to comprise two or more popul tions cannot be identified and separated by an objective and hydrologically meaningful criterion, the record shall be treated as coming from one population.

* 9. Outliers--Outliers are data points which depart significantly from the trend of the remaining data. The retention, modification, deletion of these outliers can significantly affect the statistical parameters computed from the data, especially for small samples. All procedures for treating outliers ultimately require judgment involving both mathematical and hydrologic considerations. The detection and treatment of high and low outliers are described below, and are outlined on the flow chart in Appendix 12 (figure 12-3).

If the station skew is greater than +0.4, tests for high outliers are considered first. If the station skew is less than -0.4 tests for low outliers are considered first. Where the station skew is between $\pm$ 0.4, tests for both high and low outliers should be applied before eliminating any outliers from the data set.

The following equation is used to detect high outliers:

$$X_H = \overline{X} + K_N S \tag{7}$$

where X_H = high outlier threshold in log units

$\overline{X}$ = mean logarithm of systematic peaks (X's) excluding zero flood events, peaks below gage base, and outliers previously detected.

S = standard deviation of X's

K_N = K value from Appendix 4 for sample size N

If the logarithms of peaks in a sample are greater than X_H in equation 7 then they are considered high outliers. Flood peaks considered high outliers should be compared with historic flood data and flood information at nearby sites. If information is available which indicated a high outlier(s) is the maximum in an extended period of time, the outlier(s) is treated as historic flood data as described in Section V.B.10. If useful historic information is not available to adjust for high outliers, then they should be retained as part of the systematic record. The treatment of all historic flood data and high outliers should be well documented in the analysis. *

* The following equation is used to detect low outliers:

$$X_L = \overline{X} - K_N S \qquad (8a)$$

where X_L = low outlier threshold in log units and the other terms are as defined for equation 7.

If an adjustment for historic flood data has previously been made, then the following equation is used to detect low outliers:

$$X_L = \widetilde{M} - K_H \widetilde{S} \qquad (8b)$$

where X_L = low outlier threshold in log units

K_H = K value from Appendix 4 for period used to compute $\widetilde{M}$ and $\widetilde{S}$

$\widetilde{M}$ = historically adjusted mean logarithm

$\widetilde{S}$ = historically adjusted standard deviation

If the logarithms of any annual peaks in a sample are less than X_L in equation 8a or b, then they are considered low outliers. Flood peaks considered low outliers are deleted from the record and the conditional probability adjustment described in Appendix 5 is applied.

If multiple values that have not been identified as outliers by the recommended procedure are very close to the threshold value, it may be desirable to test the sensitivity of the results to treating these values as outliers.

Use of the K values from Appendix 4 is equivalent to a one-sided test that detects outliers at the 10 percent level of significance (38). The K values are based on a normal distribution for detection of single outliers. In this Bulletin, the test is applied once and all values above the equation 7 threshold or below that from equation 8a or b are considered outliers. The selection of this outlier detection procedure was based on testing several procedures on simulated log-Pearson Type III and observed flood data and comparing results. The population skew coefficients for the simulated data were between $\pm$ 1.5, with skews for samples selected from these populations ranging between -3.67 and +3.25. The skew values * *for the observed data were between -2.19 and +2.80. Other test procedures evaluated included use of station, generalized, weighted, and zero skew. The selected procedure performed as well or better than the other procedures while at the same time being simple and easy to apply. Based on these results, this procedure is considered appropriate for use with the log-Pearson Type III distribution over the range of skews, $\pm$ 3.

10. Historic Flood Data - Information which indicates that any flood peaks which occurred before, during, or after the systematic record are maximums in an extended period of time should be used in frequency computations. Before such data are used, the reliability of the data, the peak discharge magnitude, changes in watershed conditions over the extended period of time, and the effects of these on the computed frequency curve must all be evaluated by the analyst. The adjustment described in Appendix 6 is recommended when historic data are used. The underlying assumption to this adjustment is that the data from the systematic record is representative of the intervening period between the systematic and historic record lengths. Comparison of results from systematic and historically adjusted analyses should be made.

The historic information should be used unless the comparison of the two analyses, the magnitude of the observed peaks, or other factors suggest that the historic data are not indicative of the extended record. All decisions made should be thoroughly documented. *

C. Refinements to Frequency Curve

The accuracy of flood probability estimates based upon statistical analysis of flood data deteriorates for probabilities more rare than those directly defined by the period of systematic record. This is partly because of the sampling error of the statistics from the station data and partly because the basic underlying distribution of flood data is not known exactly.

Although other procedures for estimating floods on a watershed and flood data from adjoining watersheds can sometimes be used for evaluating flood levels at high flows and rare exceedance probabilities;

procedures for doing so cannot be standardized to the same extent as the procedures discussed thus far. The purpose for which the flood frequency information is needed will determine the amount of time and effort that can justifiably be spent to obtain and make comparisons with other watersheds, and make and use flood estimates from precipitation. The remainder of the recommendations in this section are guides for use of these additional data to refine the flood frequency analysis.

The analyses to include when determining the flood magnitudes with 0.01 exceedance probability vary with length of systematic record as shown by an X in the following tabulation:

* Analyses to Include	Length of Record Available		
	10 to 24	25 to 50	50 or more
Statistical Analysis	X	X	X
Comparisons with Similar Watersheds	X	X	--
Flood Estimates from Precipitation	X	--	-- *

All types of analyses should be incorporated when defining flood magnitudes for exceedance probabilities of less than 0.01. The following sections explain how to include the various types of flood information in the analysis.

1. Comparisons with Similar Watersheds--A comparison between flood and storm records (see, e.g., (12)) and flood flow frequency analyses at nearby hydrologically similar watersheds will often aid in evaluating and interpreting both unusual flood experience and the flood frequency analysis of a given watershed. The shorter the flood record and the more unusual a given flood event, the greater will be the need for such comparisons.

Use of the weighted skew coefficient recommended by this guide is one form of regional comparison. Additional comparisons may be helpful and are described in the following paragraphs.

Several mathematical procedures have been proposed for adjusting a short record to reflect experience at a nearby long-term station. Such procedures usually yield useful results only when the gaging stations are on the same stream or in watersheds with centers not more than 50 miles apart. The recommended procedure for making such adjustments is given in Appendix 7. The use of such adjustments is confined to those situations where records are short and an improvement in accuracy of at least 10 percent can be demonstrated.

Comparisons and adjustment of a frequency curve based upon flood experience in nearby hydrologically similar watersheds can improve most flood frequency determinations. Comparisons of statistical parameters of the distribution of flows with selected exceedance probabilities can be made using prediction equations [e.g., (13), (14), (15), (16)], the index flood method (17), or simple drainage area plots. As these estimates are independent of the station analysis, a weighted average of the two estimates will be more accurate than either alone. The weight given to each estimate should be inversely proportional to its variance as described in Appendix 8. Recommendations of specific procedures for regional comparisons or for appraising the accuracy of such estimates are beyond the scope of this guide. In the absence of an accuracy appraisal, the accuracy of a regional estimate of a flood with 0.01 exceedance probability can be assumed equivalent to that from an analysis of a 10-year station record.

2. Flood Estimates from Precipitation--Floods estimated from observed or estimated precipitation (rainfall and/or snowmelt) can be used in several ways to improve definition of watershed flood potential. Such estimates, however, require a procedure (e.g., calibrated watershed model, unit hydrograph, rainfall-runoff relationships) for converting precipitation to discharge. Unless such procedures are available, considerable effort may be required to make these flood estimates. Whether or not such effort is warranted depends upon the procedures and data available and on the use to be made of the estimate.

Observed watershed precipitation can sometimes be used to estimate a missing maximum event in an incomplete flood record.

Observed watershed precipitation or precipitation observed at nearby stations in a meteorologically homogeneous region can be used to generate a synthetic record of floods for as many years as adequate precipitation records are available. Appraisal of the technique is outside the scope of this guide. Consequently, alternative procedures for making such studies, or criteria for deciding when available flood records should be extended by such procedures have not been evaluated.

Floods developed from precipitation estimates can be used to adjust frequency curves, including extrapolation beyond experienced values. Because of the many variables, no specific procedure is recommended at this time. Analysts making use of such procedures should first standardize methods for computing the flood to be used and then evaluate its probability of occurrence based upon flood and storm experience in a hydrologically and meteorologically homogeneous region. Plotting of the flood at the exceedance probability thus determined provides a guide for adjusting and extrapolating the frequency curve. Any adjustments must recognize the relative accuracy of the flood estimate and the other flood data.

VI. Reliability Application

The preceding sections have presented recommended procedures for determination of the flood frequency curve at a gaged location. When applying these curves to the solution of water resource problems, there are certain additional considerations which must be kept in mind. These are discussed in this section.

It is useful to make a distinction in hydrology between the concepts of risk and uncertainty (18).

Risk is a permanent population property of any random phenomenon such as floods. If the population distribution were known for floods, then the risk would be exactly known. The risk is stated as the probability that a specified flood magnitude will be exceeded in a specified period of years. Risk is inherent in the phenomenon itself and cannot be avoided.

Because use is made of data which are deficient, or biased, and because population properties must be estimated from these data by some technique, various errors and information losses are introduced into the flood frequency determination. Differences between the population properties and estimates of these properties derived from sample data constitute uncertainties. Risk can be decreased or minimized by various water resources developments and measures, while uncertainties can be decreased only by obtaining more or better data and by using better statistical techniques.

The following sections outline procedures to use for (a) computing confidence limits which can be used to evaluate the uncertainties inherent in the frequency determination, (b) calculating risk for specific time periods, and (c) adjusting the frequency curve to obtain the expected probability estimate. The recommendations given are guides as to how the procedures should be applied rather than instruction on when to apply them. Decisions on when to use each of the methods depend on the purpose of the estimate.

A. Confidence Limits

The user of frequency curves should be aware that the curve is only an estimate of the population curve; it is not an exact representation. A streamflow record is only a sample. How well this sample will predict the total flood experience (population) depends upon the sample size, its accuracy, and whether or not the underlying distribution is known. Confidence limits provide either a measure of the uncertainty of the estimated exceedance probability of a selected discharge or a measure of the uncertainty of the discharge at a selected exceedance probability. Confidence limits on the discharge can be computed by the procedure described in Appendix 9.

Application of confidence limits in reaching water resource planning decision depends upon the needs of the user. This discussion is presented to emphasize that the frequency curve developed using this guide is only today's best estimate of the flood frequency distribution. As more data become available, the estimate will normally be improved and the confidence limits narrowed.

B. Risk

As used in this guide, risk is defined as the probability that one or more events will exceed a given flood magnitude within a specified period of years. Accepting the flow frequency curve as accurately representing the flood exceedance probability, an estimate of risk may be computed for any selected time period. For a 1-year period the probability of exceedance, which is the reciprocal of the recurrence interval T, expresses the risk. Thus, there is a 1 percent chance that the 100-year flood will be exceeded in a given year. This statement however, ignores the considerable risk that a rare event will occur during the lifetime of a structure. The frequency curve can also be used to estimate the probability of a flood exceedance during a specified time period. For instance, there is a 50 percent chance that the flood with annual exceedance probability of 1 percent will be exceeded one or more times in the next 70 years.

Procedures for making these calculations are described in Appendix 10 and can be found in most standard hydrology texts or in (19) and (20).

C. Expected Probability

The expected probability is defined as the average of the true probabilities of all magnitude estimates for any specified flood frequency that might be made from successive samples of a specified size [(8), (21)]. It represents a measure of the central tendency of the spread between the confidence limits.

The study conducted for the Work Group (8) and summarized in Appendix 14 indicates that adjustments [(21),(22)] for the normal distribution are approximately correct for frequency curves computed using the statistical procedures described in this guide. Therefore, the committee recommends that if an expected probability adjustment is made, published adjustments applicable to the normal distribution be used. It would be the final step in the frequency analysis. It must be documented as to whether or not the expected probability adjustment is made. If curves are plotted, they must be appropriately labeled.

It should be recognized when using the expected probability adjustment that such adjustments are an attempt to incorporate the effects of uncertainty in application of the curve. The basic flood frequency curve without expected probability is the curve used in computation of confidence limits and risk and in obtaining weighted averages of independent estimates of flood frequency discharge.

The decision about use of the expected probability adjustment is a policy decision beyond the scope of this guide. It is most often used in estimates of annual flood damages and in establishing design flood criteria.

Appendix 11 provides precedures for computing the expected probability and further description of the concept.

VII. Potpourri

The following sections provide information that is of interest but not essential to use of this guide.

A. Non-conforming Special Situations

This guide describes the set of procedures recommended for defining flood potential as expressed by a flood flow frequency curve. In the Introduction the point is made that special situations may require other approaches and that in those cases where the procedures of this guide are not followed, deviations must be supported by appropriate study, including a comparison of the results obtained with those obtained using the recommended procedures.

It is not anticipated that many special situations warranting other approaches will occur. Detailed and specific recommendations on analysis are limited to the treatment of the station data including records of historic events. These procedures should be followed unless there are compelling technical reasons for departing from the guide procedures. These deviations are to be documented and supported by appropriate study, including comparison of results. The Hydrology Subcommittee asks that these situations be called to its attention for consideration in future modifications of this guide.

The map of skew (Plate I) is a generalized estimate. **Users are** encouraged to make detailed studies for their region of interest using the procedures outlined in Section V.B.3.

Major problems in flood frequency analysis at gaged locations are encountered when making flood estimates for probabilities more rare than defined by the available record. For these situations the guide described the information to incorporate in the analysis but allows considerable latitude in analysis.

B. Plotting Position

Calculations specified in this guide do not require designation of a plotting position. Section V.B.10., describing treatment of historic data, states that the results of the analysis should be shown graphically to permit an evaluation of the effect on the analysis of including historic data. The merits of alternative plotting position formulae were not studied and no recommendation is made.

A general formula for computing plotting positions (23) is

$$P = \frac{(m-a)}{(N-a-b+1)} \qquad (9)$$

where

m = the ordered sequence of flood values with the largest equal to 1

* N = number of items in data set and a and b depend upon the distribution. For symmetrical distributions a=b and the formula reduces to *

$$P = \frac{(m-a)}{(N-2a+1)} \qquad (10)$$

The Weibull plotting position in which a in equation 10 equals 0 was used to illustrate use of the historic adjustment of figure 6-3 and has been incorporated in the computer program referenced in Appendix 13, to facilitate data and analysis comparisons by the program user. This plotting position was used because it is analytically simple and intuitively easily understood (18, 24).

Weibull Plotting Position formula:

$$P = \frac{m}{N + 1} \tag{11}$$

C. Future Studies

This guide is designed to meet a current, ever-pressing demand that the Federal Government develop a coherent set of procedures for accurately defining flood potentials as needed in programs of flood damage abatement. Much additional study and data are required before the twin goals of accuracy and consistency will be obtained. It is hoped that this guide contributes to this effort by defining the essential elements of a coherent set of proedures for flood frequency determination. Although selection of the analytical procedures to be used in each step or element of the analysis has been carefully made based upon a review of the literature, the considerable practical experience of Work Group members, and special studies conducted to aid in the selection process, the need for additional studies is recognized. Following is a list of some additional needed studies identified by the Work Group.

1. Selection of distribution and fitting procedures
   (a) Continued study of alternative distributions and fitting procedures is believed warranted.
   (b) Initially the Work Group had expected to find that the proper distribution for a watershed would vary depending upon watershed and hydrometeorological conditions. Time did not permit exploration of this idea.

(c) More adequate criteria are needed for selection of a distribution.

(d) Development of techniques for evaluating homogeneity of series is needed.

2. The identification and treatment of mixed distributions.
3. The treatment of outliers both as to identification and computational procedures.
4. Alternative procedures for treating historic data.
5. More adequate computation procedures for confidence limits to the Pearson III distribution.
6. Procedures to incorporate flood estimates from precipitation into frequency analysis.
7. Guides for defining flood potentials for ungaged watersheds and watersheds with limited gaging records.
8. Guides for defining flood potentials for watersheds altered by urbanization and by reservoirs.

Appendix 1

REFERENCES

1. Brater, E. F. and J. D. Sherrill, Prediction of the Magnitudes and Frequencies of Floods in Michigan, Department of Civil Engineering, University of Michigan, Ann Arbor, 56pp., August 1971. * *

2. Gann, E. E., "Generalized Flood-Frequency Estimate for Urban Areas in Missouri," USGS Open-File Report, 18 pp., 1971.

3. Thomas, C. A., W. A. Harenberg, and J. M. Anderson, "Magnitude and Frequency of Floods in Small Drainage Basins in Idaho," USGS Water Resources Inv. 7-73, NTIS, Springfield, VA., 1973.

4. Todorovic, R. and E. Zelenhasic, "A Stochastic Model for Flood Analysis," Water Resources Research, Vol. 6, No. 6, 1970, pp. 1641-1648.

5. Carrigan, P. H., Jr., and C. S. Huzzen, "Serial Correlation of Annual Floods," International Hydrology Symposium, Fort Collins, September 1967.

6. Matalas, N. C., "Autocorrelation of Rainfall and Streamflow Minimums," U.S. Geological Survey Prof. Paper 434B, 1963.

7. Pacific Southwest Interagency Commission Report of the Hydrology Subcommittee, "Limitations in Hydrologic Data as Applied to Studies in Water Control Management," February 1966.

8. Beard, L. R., Flood Flow Frequency Techniques, Center for Research in Water Resources, The University of Texas at Austin, 1974.

9. Langbein, W. B., "Annual Floods and the Partial Duration Series" Transactions, American Geophysical Union, Vol. 30, 1949, p. 879.

10. Hardison, C. H., "Generalized Skew Coefficients of Annual Floods in the United States and Their Application" Water Resources Research, Vol. 10, No. 5, pp. 745-752.

11. U.S. Army Engineer District, Sacramento, California, Civil Works Investigations Project CW-151, Flood Volume Studies-West Coast. Research Note No. 1, "Frequency of New England Floods," July 1958.

12. Corps of Engineers, U.S. Army, "Storm Rainfall in the United States," Washington, 1945.

13. Benson, M. A., "Evaluation of Methods for Evaluating the Occurrence of Floods," USGS Water Supply Paper 1580-A, 1962.

14. Benson, M. A., "Factors Influencing the Occurrence of Floods in a Humid Region of Diverse Terrain," USGS Water Supply Paper 1580-B, 1962.

15. Benson, M. A., "Factors Affecting the Occurrence of Floods in the Southwest," USGS Water Supply Paper 1580-D, 1964.

16. Bock, P., I. Enger, G. P. Malhotra, and D. A. Chisholm, "Estimating Peak Runoff Rates from Ungaged Small Rural Watersheds," National Cooperative Highway Research Program Report 136, Highway Research Board, 1972.

17. Dalrymple, T., Flood Frequency Analyses, Manual of Hydrology: Part 3, "Flood Flow Techniques," USGS Water Supply Paper 1543-A, 1960.

18. Yevjevich, Vujica, Probability and Statistics in Hydrology, Water Resources Publications, Fort Collins, Colorado, 1972.

19. Gumbel, E. J., "The Calculated Risk in Flood Control," Applied Science Research, Section A, Vol. 5, 1955, The Hague.

20. Riggs, H. C., "Frequency of Natural Events," Journal of the Hydraulics Division, Proc. ASCE, Vol. 87, No. HY1, 1961, pp. 15-26.

21. Beard, L. R., "Probability Estimates Based on Small Normal-Distribution Samples," Journal of Geophysical Research, July 1960.

22. Hardison, C., and M. Jennings, "Bias in Computed Flood Risk," Journal of the Hydraulics Division, Proc. ASCE, Vol. 98, No. HY3, March 1972 (with discussion and closure), pp. 415-427.

23. Harter, H. L., "Some Optimization Problems in Parameter Estimation," edited by Jagdish S. Rustagi, Optimizing Methods in Statistics, Academic Press, New York, 1971, pp. 32-62.

24. Chow, V. T., Handbook of Applied Hydrology, McGraw-Hill Book Co., New York, 1964, pp. 8-28 and 8-29.

25. Hardison, C., "Accuracy of Streamflow Characteristics," USGS Professional Paper 650-D, 1969, pp. D210-D214.

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

27. McGinnis, David G., and William H. Sammons, Discussion of Paper by Payne, Neuman, and Kerri. "Daily Stream Flow Simulation," Journal of the Hydraulics Division, Proc. ASCE, Vol. 96, No. HY5, May 1970.

28. Jennings, M. E. , and M. A. Benson, "Frequency Curves for Annual Flood Series with Some Zero Events or Incomplete Data," Water Resources Research, Vol. 5, No. 1, 1969, pp. 276-280.

29. Matalas, N., and B. Jacobs, "A Correlation Procedure for Augmenting Hydrologic Data," USGS Professional Paper 434-E, 1964.

30. Gilroy, E. J. personal communication to C. Hardison, 1974.

31. Beard, L. R., Statistical Methods in Hydrology, U. S. Army Corps of Engineers, Civil Works Investigation Project CW-151, 1962.

32. Natrella, M. G., Experimental Statistics, National Bureau of Standards Handbook 91, 1963.

33. Hardison, C. H., personal communication, 1974

34. U.S. Corps of Engineers, "Regional Frequency Computation," Generalized Computer Program, The Hydrologic Engineering Center, July 1972.

35. Water Resources Council, Hydrology Committee, A Uniform Technique for Determining Flood Flow Frequencies, Bulletin 15, Washington, D.C., 1967.

36. Harter, H. L., and A. H. Moore, "A Note on Estimation from a Type I Extreme-Value Distribution," Technometrics, Vol. 9, No. 2, May 1967, pp. 325-331.

37. Thom, H. C. S., "A Note on the Gamma Distribution," Monthly Weather Review, Vol. 86, 1958, pp. 117-122.

* 38. Grubbs, Frank E. and Glenn Beck, "Extension of Sample Sizes and Percentage Points for Significance Tests of Outlying Observations," Technometrics, Vol. 14, No. 4, November 1972, pp. 847-854.

39. Tasker, Gary D., "Flood Frequency Analysis with a Generalized Skew Coefficient," Water Resources Research, Vol. 14, No. 2, April 1978, pp. 373-376.

40. Wallis, J. R., N. C. Matalas, and J. R. Slack, "Just a Moment," Water Resources Research, Vol. 10, No. 2, April 1974, pp. 211-219. *

* 41. Resnikoff, G. J., and G. J. Lieberman, <u>Tables of the Non-Central t-Distribution</u>, Stanford University Press, Stanford, California, 1957.

42. Zelen, N., and N. C. Severo, "Probability Functions," <u>Handbook of Mathematical Functions</u>, Applied Mathematics Series No. 55, U. S. National Bureau of Standards, 1904.

43. Owen, D. B., "<u>Factors for One-Sided Tolerence Limits and for Variables Sampling Plans</u>," Sandia Corporation Monograph SCR-607, March, 1963. *

Appendix 2

GLOSSARY AND NOTATION

Glossary

The terms used in this guide include definitions taken from references listed in the Bibliography or from "Nomenclature for Hydraulics," Manual 43, American Society of Civil Engineers, 1962, and from definitions especially prepared for this guide. For more technical definitions of statistical terms, see "Dictionary of Statistical Terms" by M. G. Kendall and W. R. Buckland, Hafner Publishing Company, New York, 1957.

TERM	Definition
Annual Flood	The maximum momentary peak discharge in each year of record. (Sometimes the maximum mean daily discharge is used.)
Annual Flood Series	A list of annual floods.
Annual Series	A general term for a set of any kind of data in which each item is the maximum or minimum in a year.
Array	A list of data in order of magnitude; in flood-frequency analysis it is customary to list the largest value first, in a low-flow frequency analysis the smallest first.
Broken Record	A systematic record which is divided into separate continuous segments because of deliberate discontinuation of recording for significant periods of time.

Coefficient of Skewness — A numerical measure or index of the lack of symmetry in a frequency distribution. Function of the third moment of magnitudes about their mean, a measure of asymmetry. Also called "coefficient of skew" or "skew coefficient."

Confidence Limits — Computed values on both sides of an estimate of a parameter that show for a specified probability the range in which the true value of the parameter lies.

Distribution — Function describing the relative frequency with which events of various magnitudes occur.

Distribution-Free — Requiring no assumptions about the kind of probability distribution a set of data may have.

Exceedance Frequency — The percentage of values that exceed a specified magnitude, 100 times exceedance probability.

Exceedance Probability — Probability that a random event will exceed a specified magnitude in a given time period, usually one year unless otherwise indicated.

Expected Probability — The average of the true probabilities of all magnitude estimates for any specified flood frequency that might be made from successive samples of a specified size.

Generalized Skew Coefficient — A skew coefficient derived by a procedure which integrates values obtained at many locations.

Homogeneity — Records from the same populations.

Incomplete Record — A streamflow record in which some peak flows are missing because they were too low or high to record or the gage was out of operation for a short period because of flooding.

Level of Significance — The probability of rejecting a hypothesis when it is in fact true. At a "10-percent" level of significance the probability is 1/10.

* *Mean-Square Error* — Sum of the squared differences between the true and estimated values of a quantity divided by the number of observations. It can also be defined as the bias squared plus the variance of the quantity. *

Method of Moments — A standard statistical computation for estimating the moment of a distribution from the data of a sample.

Nonparametric — The same as distribution-free.

Normal Distribution — A probability distribution that is symmetrical about the mean, median, and mode (bell-shaped). It is the most studied distribution in statistics, even though most data are not exactly normally distributed, because of its value in theoretical work and because many other distributions can be transformed into normal. It is also known as Gaussian, The Laplacean, The Gauss-Laplace, or the Laplace-Gauss distribution, or the Second Law of Laplace.

Outlier — Outliers (extreme events) are data points which depart from the trend of the rest of data.

Parameter — A characteristic descriptor, such as a mean or standard deviation.

Percent Chance — A probability multiplied by 100.

Population — The entire (usually infinite) number of data from which a sample is taken or collected. The total number of past, present, and future floods at a location on a river is the population of floods for that location even if the floods are not measured or recorded.

Recurrence Interval (Return Period, Exceedance Interval) — The average time interval between actual occurrences of a hydrological event of a given or greater magnitude. In an annual flood series, the average interval in which a flood of a given size is exceeded as an annual maximum. In a partial duration series, the average interval between floods of a given size, regardless of their relationship to the year or any other period of time. The distinction holds even though for large floods recurrence intervals are nearly the same for both series.

Sample — An element, part, or fragment of a "population." Every hydrologic record is a sample of a much longer record.

Skew Coefficient — See "coefficient of skewness."

Standard Deviation

A measure of the dispersion or precision of a series of statistical values such as precipitation or streamflow. It is the square root of the sum of squares of the deviations from the arithmetic mean divided by the number of values or events in the series. It is now standard practice in statistics to divide by the number of values minus one in order to get an unbiased estimate of the variance from the sample data.

Standard Error

An estimate of the standard deviation of a statistic. Often calculated from a single set of observations. Calculated like the standard deviation but differing from it in meaning.

Student's t Distribution (t-distribution)

A distribution used in evaluation of variables which involve sample standard deviation rather than population standard deviation.

Test of Significance

A test made to learn the probability that a result is accidential or that a result differs from another result. For all the many types of tests there are standard formulas and tables. In making a test it is necessary to choose a "level of significance," the choice being arbitrary but generally not less than the low level of 10 percent nor more than the high level of 1 percent.

Transformation — The change of numerical values of data to make later computations easier, to linearize a plot or to normalize a skewed distribution by making it more nearly a normal distribution. The most common transformations are those changing ordinary numerical values into their logarithms, square roots or cube roots; many others are possible.

Variance — A measure of the amount of spread or dispersion of a set of values around their mean, obtained by calculating the mean value of the squares of the deviations from the mean, and hence equal to the square of the standard deviation.

Weighted Means — A value obtained by multiplying each of a series of values by its assigned weight and dividing the sum of those products by the sum of the weights.

Notation

Appendix notation is described in each Appendix. While most notation is consistent, slight variations do occur.

Notation	Explanation
* A	Fitting parameter used in equation 6.
a	Variate in equations 9 and 10 which depends upon the distribution (23).
* B	Fitting parameter used in equation 6.
b	Variate in equation 9 which depends upon the distribution (23)
G	Skew coefficient of logarithms of annual peak discharges
$\bar{G}$	Generalized skew coefficient
* $\tilde{G}$	Historically adjusted skew coefficient
G_W	Weighted skew coefficient
H	Historic record length
K_H	K value from Appendix 4 for historic period H
K	Pearson Type III deviate
* K_N	K value from Appendix 4 for sample size N
$\tilde{M}$	Historically adjusted mean logarithm
MSE	Mean-square error
$MSE_{\bar{G}}$	Mean-square error of generalized skew
MSE_G	Mean-square error of station skew
m	Ordered sequence of flood values, with the largest equal to 1
N	Number of items in data set
P	Exceedance probability
Q	Peak discharge, cfs
S	Standard deviation of logarithms of annual peak discharges
* $\tilde{S}$	Historically adjusted standard deviation

Notation

SE_G Standard error of sample skew coefficient, which for samples from a normal distribution can be estimated as:

$$SE_G = \sqrt{\frac{6N(N-1)}{(N-2)(N+1)(N+3)}}$$

SE_S Standard error of sample standard deviation, can be estimated as:

$$SE_S = \frac{S\sqrt{1 + 0.75\ G^2}}{\sqrt{2N}}$$

$SE_{\overline{X}}$ Standard error of sample mean, can be estimated as:

$$SE_{\overline{X}} = \frac{S}{\sqrt{N}}$$

T Recurrence interval in years

X Logarithm of peak flow

$\overline{X}$ Mean logarithm of peak flows

* X_H High outlier threshold in log units *

X_L Low outlier threshold in log units