Flood-Frequency Analyses

Manual of Hydrology: Part 3. Flood-Flow Techniques

GEOLOGICAL SURVEY WATER-SUPPLY PAPER 1543-A

Flood-Frequency Analyses

By TATE DALRYMPLE

Manual of Hydrology: Part 3. Flood-Flow Techniques

GEOLOGICAL SURVEY WATER-SUPPLY PAPER 1543-A

Methods and practices of the Geological Survey

UNITED STATES GOVERNMENT PRINTING OFFICE, WASHINGTON : 1960

UNITED STATES DEPARTMENT OF THE INTERIOR

FRED A. SEATON, *Secretary*

GEOLOGICAL SURVEY

Thomas B. Nolan, *Director*

For sale by the Superintendent of Documents, U.S Government Printing Office
Washington 25, D.C.

CONTENTS

	Page
Abstract	1
Introduction	1
Flood frequency at a gaging station	5
Two kinds of flood series	5
Annual floods	5
Partial-duration series	6
Relation between the two series	6
Flood-frequency curves	7
Discharge curves	7
Stage curves	7
Flood-volume curves	8
Listing of data	8
Annual-flood peaks	8
Partial-duration series peaks	10
Flood volumes	12
Historical data	13
Plotting positions	14
Methods	15
Plotting historical data	16
Plotting paper	19
Fitting frequency graphs	20
Single-station analysis	21
Regional flood frequency	25
Basic frequency curve	27
Homogeneity test	27
Mean annual flood	29
Physiographic factors	29
Meteorologic factors	30
Composite factors	30
Computation procedure	31
Selection of stations	31
Selection of base period	32
Adjustment of records to base period	34
Preliminary frequency curve	37
The mean annual flood	37
Homogeneity test	38
Computation of median flood ratios	39
Definition of regional frequency curve	40
Estimation of mean annual flood	41
Definition of a frequency curve	45
Summary of procedure	45
Special treatment for large streams	46

Page

Plotting positions in frequency analysis, by W. B. Langbein ... 48
  Median ... 49
  Mean ... 49
  Mode ... 50
Characteristics of frequency curves based on a theoretical 1,000-year record, by M. A. Benson ... 51
  Base data ... 51
  Mean annual flood ... 54
  Short periods of record ... 54
    Periods used ... 54
    Drawing of frequency curves ... 54
    Arithmetic versus graphical mean annual flood ... 56
    Reliability of mean annual-flood values ... 61
    "Working range" of mean annual floods ... 61
    Reliability of flood magnitudes ... 63
  Composite flood-frequency curves ... 65
    Periods combined ... 65
    Results from composite graphs ... 67
    Effect of number of stations combined ... 67
    Median versus average ratios ... 71
    Theoretical distribution of composite curves ... 71
  Conclusions ... 73
Selected references ... 74
Index ... 79

ILLUSTRATIONS

FIGURE 1. Flood data listing ... 9
2. Tabulation of flood-volume data ... 13
3. Flood-frequency curve, rectangular scale ... 22
4. Flood-frequency curve, logarithmic scale ... 23
5. Partial-duration series flood-frequency curve ... 24
6. Annual flood volume-frequency curves ... 25
7. Youghiogheny and Kiskiminetas River basins, showing location of gaging stations ... 33
8. Bar graph showing period of record of maximum annual peaks at gaging stations ... 34
9. Tabulation of flood data, Big Piney Run near Salisbury, Pa ... 35
10. Correlation of annual floods, Big Piney Run versus Stony Creek ... 36
11. Preliminary frequency curve, Big Piney Run near Salisbury, Pa ... 37
12. Homogeneity test chart ... 39
13. Regional frequency curve ... 41
14. Variation of mean annual flood with drainage area ... 42
15. Determination of mean annual flood from a short record ... 44
16. Flood-frequency curve for a 100-square-mile drainage basin ... 45
17. Variation of mean annual flood with distance along stream ... 46
18. Relation of flood frequencies to drainage area ... 47

		Page
FIGURE	19. Theoretical frequency curve	53
	20. Frequency curves for 10-year periods	57
	21. Frequency curves for 25-year periods	58
	22. Frequency curves for 50-year periods	59
	23. Frequency curves for 100-year periods	60
	24. Recurrence intervals of 2.33-year flood	62
	25. Computing range for mean annual flood	63
	26. Recurrence intervals of 10-year flood	66
	27. Composite frequency curves based on ten 10-year periods	68
	28. Composite frequency curves based on ten 25-year periods	69
	29. Composite frequency curves based on ten 50-year periods	70
	30. Theoretical distribution of composite curves	72

MANUAL OF HYDROLOGY: PART 3. FLOOD-FLOW TECHNIQUES

FLOOD-FREQUENCY ANALYSES

By Tate Dalrymple

ABSTRACT

This report describes the method used by the U.S. Geological Survey to determine the magnitude and frequency of momentary peak discharges at any place on a stream, whether a gaging-station record is available or not. The method is applicable to a region of any size, as a river basin or a State, so long as the region is hydrologically homogeneous.

The analysis provides two curves. The first expresses the flood discharge-time relation, showing variation of peak discharge, expressed as a ratio to the mean annual flood, with recurrence interval. The second relates the mean annual flood to the size of drainage area alone, or to the size area and other significant basin characteristics.

A frequency curve may be defined for any place in the region by use of these two curves. The procedure is: (a) measure the drainage area and other appropriate basin characteristics from maps; (b) from the second curve, select the mean annual flood corresponding to the proper drainage area factors; (c) from the first curve, select ratios of peak discharge to mean annual flood for selected recurrence intervals, as 2, 10, 25, and 50 years; and (d) multiply these ratios by the mean annual flood and plot the resulting discharges of known frequency to define the frequency curve.

Two reports not previously given general circulation are included as sections of this report. These are "Plotting Positions in Frequency Analysis" by W. B. Langbein, and "Characteristics of Frequency Curves Based on a Theoretical 1,000-Year Record" by M. A. Benson.

INTRODUCTION

This report on flood-frequency analyses reflects work done by many investigators. Of the Geological Survey investigators, special acknowledgment is due to W. B. Langbein, who has made many contributions to the techniques described. Acknowledgment is due also to M. A. Benson of the Washington office staff, and to many others in the field offices who helped prepare about 25 regional studies for all parts of the United States.

Techniques used by the Surface Water Branch, Water Resources Division, of the Geological Survey in making flood-frequency studies are described. The report incorporates material previously released for official use only, as part of the Handbook for Hydrologists, chapter 1, "Instructions for Flood Frequency Compilations," revised October 1949, and chapter 8, "Flood Frequency Analyses," revised May 1950.

Knowledge of the magnitude and probable frequency of recurrence of floods is necessary to the proper design and location of structures such as dams, bridges, culverts, levees, highways, waterworks, sewage-disposal plants, and industrial buildings. An engineer often must design a structure which may be damaged or destroyed by occasional floods of varying magnitude. The frequency with which such damage may occur must be considered in determining the size or strength of the structure, its location, or the feasibility of building it at all. The problem is an economic one, involving computation of the total annual cost of maintaining a structure of a given design compared to the cost for other designs.

Either overdesign or underdesign of structures involves excessive costs on a long-time basis. The initial cost of a bridge that is designed to pass only a 5-year flood may be small, but the cost of rebuilding it at an average 5-year interval would be large. A bridge at the same site which is built to pass the 100-year flood might be extremely costly. Intermediate design would provide a bridge at that site for the lowest average annual cost.

Knowledge of flood frequency is necessary also to flood insurance and flood zoning, activities which are now considered on a broad scale. Without such knowledge, these activities would be seriously handicapped if not prohibited. Problems of flood insurance and flood zoning are economically important and they make the development of sound flood-frequency methods imperative.

The use of the flood-frequency method has met some criticism, largely because it has been abused. The method has little place in determining maximum limits of flood design, that is, "the maximum possible flood." With the ordinary streamflow record of 25-year length, errors of sampling introduce large errors in judging the magnitude of the greater floods. However, if properly computed and conservatively interpreted, flood-frequency analysis is a valuable hydrologic tool. The subject attracts many students of hydrology and it has benefited by their extensive writings.

Foster (1924) described the application of frequency curves to engineering problems of the flexible skew-distribution curves devised by Karl Pearson for frequency distributions of annual floods. The fitting of these curves requires selection of type and calculation of mean and coefficients of variation and skew. Fisher and Tippett (1928) developed frequency distributions of maximum values, subsequently applied by Gumbel (1945a) to floods. Hazen (1930) published a general treatise on the determination of frequency and magnitude by the use of logarithmic skew-frequency curves.

Jarvis (1936) edited a comprehensive source book on flood frequency containing chapters by Saville (p. 398–420) on the methods of Fuller, Foster, Hazen, Goodrich, and Slade; by Slade (p. 421–432) giving an analysis of the errors inherent in calculations of the mean, coefficient of variation, and the coefficient of skew with small samples; by Horton (p. 433–450) presenting his integral-frequency formula, in which flood magnitudes always continue to increase as the recurrence interval increases, but they increase towards a finite limit, not towards infinity; and by Bernard (p. 451–461), who discusses the determination of floodflow by the unit-hydrograph method.

Gumbel (1941) presented a paper on the return period of floodflows. Powell (1943) introduced the Gumbel method to engineers. Kinnison and Colby (1945) correlated peak discharges with measurable drainage-basin characteristics.

The viewpoints and theories expressed by these writers, although instructive, have not always been consistent. Knowledge gained from a study of these works of statisticians, hydrologists, and engineers, plus what has been learned in the preparation of 24 statewide or regional flood-frequency reports, has led to a method of analysis that is presented in this report. The method reflects the latest developments based on a continuing study of the subject by engineers of the Water Resources Division of the Geological Survey. The method has been revised several times in recent years and probably will be revised in the future.

The discharge records collected by the Geological Survey and other Federal, State, and private agencies at about 7,000 places in the United States are the records upon which flood-frequency studies are based. Most of the records are short and the sampling errors are correspondingly large, or the records for different streams are for different periods of time. Few records are for as long as 60 years, and most are for less than 30 years. It is doubtful that a rigid mathematical treatment is justified for such short records.

An important element in statistical analysis is the skewness of the data; Slade (1936, p. 426) has aptly remarked "* * * that skewness is never a truly significant characteristic when the sample from which it is computed has less than about 140 items * * * and it is quite meaningless to use this measure when there are 50 or fewer items." The hopelessness of obtaining excellent results by rigid application of complex statistical analysis is obvious. For Geological Survey reports only graphical definition of the frequency curve is contemplated.

540570 O - 60 - 2

A curve based on a gaging-station record applies only to the site of the station; generally the information is wanted for an ungaged point. Investigations have been made of the possibilities of combining the flood data for all gaging stations in a drainage basin or a larger region, and of relating the resulting flood-frequency function to measurable characteristics of the drainage basin. This procedure would, in the first instance, reduce the larger sampling errors and in the second instance, give the data regional significance and so make the flood-frequency studies applicable to ungaged areas. In general, concern is with floods not exceeding 50- to 100-year recurrence intervals, as these satisfy the needs of most engineering studies.

The ultimate objective of the analysis generally is to prepare a regional flood-frequency report. First, an analysis must be made of each gaging-station record, resulting in a frequency curve for the station, and second, these station-frequency curves must be combined to give results applicable to any stream in a region. The need is for as many gaging-station records as possible, and each for as long a period of time as possible. Records used in the analysis should not be appreciably affected by works of man. The effect on floodflows by works of man should be studied independently, and these effects added to natural-flow frequency relations.

The regional flood-frequency method developed by the Geological Survey provides two basic curves. The first expresses the flood discharge-time relation, showing variation of peak discharge, expressed as a ratio to the mean annual flood, with recurrence interval. The second relates the mean annual flood of the size of the drainage area alone, or to the size and other significant basin characteristics. A frequency curve may be defined for any place in the region by use of these two curves.

Significant features of the flood-frequency method described in this report are as follows:

1. It is concerned with momentary peak discharges.
2. Recurrence intervals are computed by the formula

$$T=\frac{(n+1)}{m}$$

3. Curves are fitted graphically.
4. The mean annual flood is defined as the flood having a recurrence interval of 2.33 years.
5. A means is provided for computing flood frequencies of natural flow on any stream, gaged or ungaged, in a region.

In addition to describing the Geological Survey's flood-frequency method, this report includes two sections on reports by Survey authors on related subjects; these reports have not been published previously in a generally distributed journal. One of these reports, "Plotting Positions in Frequency Analysis," prepared in November 1954, by W. B. Langbein, gives a derivation of the formula adopted for computing plotting positions. The other report, "Characteristics of Frequency Curves Based on a Theoretical 1,000-Year Record" by M. A. Benson, furnishes a measure of the reliability of the adopted flood-frequency method.

FLOOD FREQUENCY AT A GAGING STATION

An analysis must be made of each gaging-station record before a regional study can be started. Three kinds of flood-frequency curves may be prepared: (a) discharge; (b) stage; and (c) volume. Discharge-frequency curves are the most common; these are the basis for the most often desired regional reports. Stage-frequency curves apply only to one gage and therefore are not suitable for regional treatment; these are basic to flood-insurance and flood-zoning problems. Volume-frequency curves, with proper caution, may be regionalized, and are useful in the study of water supply and storage requirements. This report is concerned mostly with discharge-frequency analysis, although the same techniques may be applicable to stage and volume analyses.

The time scale for frequency curves is the recurrence interval. This term is defined as the average interval of time within which a flood of a given magnitude will be equaled or exceeded once. A flood having a recurrence interval of 10 years is one that has a 10 percent chance of recurring in any year; likewise a 50-year flood has a 2 percent chance, and a 100-year flood has a 1 percent chance of recurring in any year.

TWO KINDS OF FLOOD SERIES

Two methods of treating flood data for studying the frequency of floods are in common use. The first is the annual-flood array and the second is the partial-duration series. Although most analysts take a tolerant view, a few are active protagonists for one method as against the other. The differences are largely a matter of definition.

ANNUAL FLOODS

An annual flood is defined as the highest mometary peak discharge in a water year. The use of only one flood in each year is the most frequent objection to the use of annual floods. Infrequently, the second highest flood in a given year, which is omitted in the above definition, may outrank many annual floods.

PARTIAL-DURATION SERIES

The objection noted under annual floods is resolved by listing all floods that are greater than a selected base without regard to number within any given time period. The base is generally selected as equal to the lowest annual flood so that at least one flood in each year is included, however, in a long record, the base is generally raised so that on the average only 3 or 4 floods a year are included. The only other criterion followed in selecting the floods is that each peak be individual; that is, be separated by substantial recession in stage and discharge.

An objection to the use of the partial-flood series is that the floods listed may not be fully independent events; closely consecutive flood peaks may actually be one flood.

The greater number of floods listed in the partial-duration series might be an advantage, particularly if the record is short. However, most of the additional floods are of low discharge and plot where the curve is well defined; the high-discharge floods are generally identical with those in the annual-flood series.

RELATION BETWEEN THE TWO SERIES

A definite relation between values in the two series exists (Langbein, 1949; Chow, 1950). The following table shows comparative values of recurrence intervals by the two methods:

Recurrence intervals in years

Partial-duration series	*Annual floods*
0.5	1.16
1.0	1.58
1.45	2.00
2.0	2.54
5.0	5.52
10	10.5
20	20.5
50	50.5
100	100.5

There is distinction in meaning between the recurrence interval of these floods. In the annual-flood series the recurrence interval is the average interval in which a flood of given size will recur as an annual maximum. In the partial-duration series, the recurrence interval is the average interval between floods of a given size regardless of their relation to the year or any other period of time. This distinction remains, although for large floods the two approach numerical equality.

The simplicity of the annual flood method is an attractive statistical feature, and this method is used by the Geological Survey. Where a frequency curve derived from the partial-duration series is desired, an annual flood curve is prepared and converted to the partial-duration series by the relation expressed in the preceding table.

FLOOD-FREQUENCY CURVES

Three kinds of flood-frequency curves may be prepared from gaging-station records. These are: (a) discharge curves, derived from momentary peak rates of flow; (b) stage curves, relating momentary peak stages to time; and (c) volume curves, relating time to the maximum 1-day, 2-day, 5-day, 10-day, or other-day, discharge.

DISCHARGE CURVES

Discharge-frequency curves are the most common and their preparation will be described in detail. Discharge curves derived from gaging-station records are the basis for establishing regional relation; this subject will be treated in another section of this report.

STAGE CURVES

Where it is necessary to compute frequencies of stage occurrence for purposes of prediction, thought must be given to the nature of the stage-discharge relation. If the stage-discharge relation has remained virtually stable throughout the period of record, frequencies can either be computed directly from stages, or can be computed from discharges first, then transformed to stages by means of the stage-discharge relation. If there has been shifting of the stage-discharge relation, past stages may have little connection with expected stages. In such case it may be best to assume the most recent stage-discharge relation will hold in the future. After frequencies for discharge are computed, they can be transformed to stage by means of the most recent rating. Sandy channels may shift frequently. If the shifting is entirely random, it is necessary to work on the basis of past stages alone. However, if such shifting is in one direction, either a successive aggradation or degradation of the channel, then the best approach is to compute first the discharge frequencies, and to transform to stage by means of an assumed stage-discharge relation based on the previous trend. Frequencies based on such assumptions would be valid only for a definite period of time, and would have to be revised periodically.

Ice-affected streams involve other problems yet unsolved. If peaks due to ice jams are among the events included in an array of high stages, it is not possible to transform from discharge to stage frequencies. If a stream has either a constant stage-discharge relation or frequent random shifts, a stage-frequency study can be made based on stages alone. Where the stage-discharge relation is not constant, but shifting either progressively or at infrequent intervals, some method must be devised which treats the effects of ice jams separately, then combines the resultant frequencies.

FLOOD-VOLUME CURVES

There are certain problems in which flood volumes are of as great concern as flood peaks. These problems involve reservoir design, where it is desired to provide a measure of flood-control storage. The frequency of flood volume can be determined by the same method as the frequency of flood peaks. An example is given using the anual-flood method.

LISTING OF DATA

A form (Survey form 9–179) has been prepared for listing data, and is reproduced as figure 1. Columns 1–4 of this form should be filled in when data are listed. Column 1 is the water year; list the year as “1956” not “1955–56” for the year ending September 30, 1956. Show dates in column 2 as “Nov. 14, 1956” or “June 19, 1957”; and show the calendar year to avoid any chance for mistake when recording peaks occurring during months of October, November, and December. List the gage height in column 3 and the discharge in column 4; a dash in the discharge column, as shown for the February 26, 1941 peak, indicates the discharge was less than the base.

A horizontal line should be shown across: (a) column 1 if there is a break in the record; (b) column 3 where there has been a change in datum and gage heights have not been adjusted to one datum; (c) column 4 if a change in location of gage has affected the stage-discharge relation; and (d) columns 1–4 where both gage heights and discharges are not exactly comparable, as when records for two adjacent stations are combined.

ANNUAL-FLOOD PEAKS

List all complete years of record. No selection should be made of part of a record to be used except to the extent of using the largest continuous period or using great historical floods.

If an annual peak may have occurred while the recorder was out of operation, make an estimate of the peak discharge, if at all possible. Do not omit any years during the period of record.

A record may begin in April a few days before a large flood which is not exceeded for the remainder of the year and examination of adjacent station records indicates that there was little flood activity prior to the April flood. The recorded flood may then be accepted as an annual flood.

Fragmentary historical flood data should be listed. Data of this kind may prove to be valuable, and should be obtained wherever possible.

9-179
(October 1952)
FLOOD DATA

UNITED STATES DEPARTMENT OF THE INTERIOR
GEOLOGICAL SURVEY
WATER RESOURCES DIVISION

File ....

Flood data for Blue River at Richmond, Mo.
Drainage area 508 square miles. Period of record 1938-57
Flood data for momentary peak discharges greater than 3,500 cfs.

Water year	Date	Gage height (Feet)	Discharge: CFS	Discharge: Ratio to $Q_{2.33}$	Annual floods: Order (M)	Annual floods: Recurrence interval (Years)	Partial duration series: Order (M)	Partial duration series: Recurrence interval (Years)	Remarks
1	2	3	4	5	6	7	8	9	10
1876	Sept. 16, 1876	21.2	21,500		—	83.0			Highest prior to 1938
1938	Aug. 22, 1938	8.20	2,920		19	1.11			
1939	May 14, 1939	3.60	500		20	1.05			
1940	June 1, 1940	19.62	10,300		7	3.00	6	3.17	
	July 18, 1940	15.45	4,470				41	.46	
1941	Feb. 26, 1941	17.95	—						Ice jam
	Sept. 5, 1941	17.55	6,330		12	1.75	21	.90	
1942	Mar. 4, 1942	15.50	4,530		17	1.24	38	.50	
	July 19, 1942	14.22	3,840				44	.43	
1943	June 1, 1943	16.13	4,900		16	1.31	32	.59	
1944	Oct. 13, 1943	20.38	13,600		2	10.50	2	9.50	
	May 21, 1944	16.42	5,110				31	.61	
	July 4, 1944	19.57	9,600				8	2.38	
1945	May 8, 1945	13.88	3,640		18	1.17	47	.40	
1955	Oct. 7, 1954	14.62	3,540				49	.39	
	Nov. 2, 1954	19.23	8,280		9	2.33	11	1.73	
	June 20, 1955	17.37	5,470				29	.66	
1956	Mar. 15, 1956	16.62	4,750				34	.56	
	Apr. 11, 1956	14.37	3,510				51	.37	
	Apr. 26, 1956	18.94	7,680				13	1.46	
	May 14, 1956	19.12	8,080		10	2.10	12	1.58	
1957	Nov. 12, 1956	17.30	6,300		14	1.50	23	.83	
	Mar. 13, 1957	13.82	3,740				46	.41	
	Apr. 22, 1957	13.38	3,520				50	.38	
	May 23, 1957	16.54	5,560				27	.70	

Graphical mean annual flood ($Q_{2.33}$) .... cfs for period ....

Sheet 2 of 2 Listed by T.D. Date 6/27/58 Checked by CHH Date 6-28-58

U. S. GOVERNMENT PRINTING OFFICE 16—67884-2

FIGURE 1.—Flood data listing.

Floods are to be listed in chronologic order. Peak stages and discharges are both to be listed. The peak stages should be those of the stream, that is, on outside gage and not necessarily on inside gage unless data show the stages to be reasonably the same. Peaks only are included in this table. Daily discharges are excluded although previous reports (Jarvis, 1936) show daily discharges almost exclusively; that situation probably reflected the practice adopted by private practitioners to whom only daily discharges were available.

Special effort should be made to determine flood peaks from the original observations or graphs except possibly on large rivers where the daily discharges and the peaks are about the same.

Stages are included for their own intrinsic value and to enable comparison with discharges a test of consistency of the record. Flood stages in many places are more useful than discharges, especially where it is desirable to place valuable property above flood levels of specified recurrence interval.

For streams with loop ratings, or those subject to rate of change, or backwater or ice effects, peak stages are not concurrent with peak discharges. For such streams peak stages should be listed independently of peak discharges. For areas where ice is a factor add a column showing ice effect in feet, for each peak listed (Prior, 1949, p. 12).

List the highest observed peak in each water year in chronologic order. The list of annual flood peaks will not necessarily be the same as the yearly maximums published in the annual water-supply papers. Where maximums are shown which occur near the end or the beginning of a water year, examine them, first to see whether they represent peaks, and second to see that they are included only once in the list of peaks.

For example, on the Suwanee River at White Springs, Fla., the maximum discharge for the 1907 water year was not a peak—it occurred on a rising stage at 12 p.m., September 30. The peak of this rise came 4 days later on October 4. Use the highest peak during the year, that of August 24–25, 1907, in the list of annual flood peaks, although the discharge is less than the maximum yearly. The peak of October 4 is surpassed by that of January 1–2 in the 1908 water year, and so is not included. For this same station, the maximum discharge for both the 1928 and the 1929 water years occurred on the same rise, the peak of which lasted through September 30 and October 1, 1928, at the turn of the water year. In a listing of annual peak discharges, do not include this peak twice. The peak for April 30, 1928, which is the largest peak, should be included except the one for September 30–October 1, in the 2 water years, and exclude a listing of that for September 30, 1928. The September 30–October 1 peak is assigned to the 1929 water year to include the April 30, 1928 peak which was the second highest in the 2 water years.

PARTIAL-DURATION SERIES PEAKS

For partial-duration series, list all peaks in which the discharge exceeds a chosen base discharge, regardless of the number of peaks occurring in a year. Peak discharges, other than the annual maximum, for many gaging stations are published in the annual series of water-supply papers that present gaging-station data. The base for listing flood peaks for the frequency-report tabulations should be the same as that for listing flood peaks in the annual reports. Rules taken from instructions for preparation of annual reports are given below:

1. Generally, list peaks for all stations for which a recording-gage record is available.
2. List peaks for nonrecording-gage stations if the frequency of gage readings is such that reasonably accurate graphs can be drawn.
3. Do not list peaks for canals, ditches, drains, or any stream where the peaks are subject to substantial control by man. For stations that are affected by a few or minor manmade regulation, either list all the peaks, regardless of whether some may be affected by regulation, or list none at all.
4. Omit peaks for stations where the record is incomplete and where peaks above the base may have occurred during the period of nonoperation. If, however, the period of nonoperation was one of low flow with no possibility of any floods, the list of peaks can be shown exactly like any complete-year station.
5. Omit peaks for streams where the crests are so flat that the peaks generally are either the same as the daily mean discharge or exceed it by such a small amount that the daily mean discharge for all practical purposes is the same as the peak. If the peaks, in general, do not exceed the daily mean by more than 5 percent, list none although an occasional peak may substantially exceed the daily mean.

For those stations for which peak discharge will be published, the following rules should be used for determining which peaks to publish and in what form to publish them:

1. Publish all peaks for which discharge equals or exceeds a chosen base discharge, regardless of the number of peaks occurring in a year. The gage height corresponding to the peak discharge will also be given. The number of peaks should average about three per year if the base is chosen properly. (Suggestions for selecting the base discharge are given on page 12.)
2. Indicate the selected base in the manuscript and in the published reports each year so that the user of the record will know what the base is at each station for which peaks are published.
3. Publish only the highest peak of 2 or more occurring within 48 hours of each other, unless it is probable that the peaks in that period are independent, a condition which will occur at times on some streams. The peaks are independent if the hydrograph recedes to well-defined troughs as defined in paragraph

4. List only the first, if 2 dependent peaks occurring within 48 hours of each other happen to be equal.
4. Do not publish a peak unless the discharge of the trough between it and the adjacent higher peak is 25 percent or more below the discharge of the lower peak.
5. Publish for periods of diurnal peaks caused by snowmelt, only the highest occurring during each distinct period of melting regardless of the fact that other peaks may fulfill the preceding requirements.
6. List all maximum annual peaks, although they may be less than the base discharge.
7. At recording-gage stations peaks may occur while the recorder is out of operation, also peaks sometimes occur during ice-affected periods. It would be misleading to exclude mention of peaks that otherwise fit the criteria because complete data are lacking. In such cases, estimates must be made or the whole year eliminated.

Criteria for selecting base discharge.—The following instructions for selecting the base discharge are followed in the preparation of annual reports.

1. For stations for which flood-frequency data have been compiled, use the same base if it has been found satisfactory. A discharge that is exceeded on an average of three times per year is low enough.
2. For stations with records of more than 5 years but for which no flood-frequency data have been compiled, list the annual flood peaks, compute their recurrence interval, and choose as the base a discharge whose recurrence interval is 1.15 years.
3. For stations with records of 5 years or less, select a base guided by judgment and by comparison with other stations. The base selected can be modified as more data become available. A base should be selected that is a little low so that if it does become necessary to change the base, a higher one can be selected.
4. Changes in the base are not desirable; after once selecting a base retain it unless it proves entirely unsatisfactory. A revision upward in base does not affect the continuity of the array of peaks above the new base. However, a revision downward in the base means that all the lists of peaks above the higher base as previously published are incomplete so far as the new base is concerned.

FLOOD VOLUMES

Flood-volume data may be listed on a standard computation form (Survey form 9–230) or other suitable form. Separate lists should be prepared of the maximum 1-day discharge, maximum 2-day (con-

9-230

UNITED STATES
DEPARTMENT OF THE INTERIOR
GEOLOGICAL SURVEY

Flood volumes of Licking River at Toboso, Ohio

Water Year	Peak dischrge	Total runoff on consecutive days: 1 day	2	3	4	5	7	10	15
				(cfs – days)					
1922	16,300	13700	20,400	23680	26680	28,520	30920	34785	40175
23	4,720	4,720	7,170	8930	10,130	11,330	13,810	15,958	24,178
24	20,000	13400	23,140	26,540	28,550	32,210	38,550	45,610	51,208
25	4,660	4660	7,210	9,270	11,160	12,082	13,079	14,196	16,584
26	9,380	8,180	13,730	16,700	18,380	19,560	21,520	25,650	38,260
27	16,800	13,600	19,800	25,780	32,310	37,190	41,850	45,660	57,190
28	15,000	13400	19,600	25,250	28,510	31,390	35,970	38,820	43,957
29	23,600	15,700	30,500	34,960	37,840	42,200	45,640	53,530	58,968
1930	9,740	8,030	15,400	20,450	22,970	25,300	37,690	46,640	58,490
31	4,460	3,600	5030	5,848	6,656	7,222	8,000	9,003	10,263
32	10,100	7,750	11,960	17,730	23,430	25,970	29,350	41,500	51,650
33	13,000	11,600	18,940	24,690	31,780	35,690	41,600	48,630	57,450
34	6,090	4780	8,390	9,930	10,980	11,748	12,490	13,220	14,599
35	20,000	15,400	23,770	28,290	30,650	34,520	37,380	40,850	49,110
36	9,740	9,020	16,630	24,420	27,270	29,060	31,440	34,394	38,127
37	25,000	21,700	31,620	41,800	60,120	70,120	83,880	102,280	135,570
38	12,400	10,500	17,890	26,600	31,080	33,610	37,300	41,295	52,245
39	13,100	11,700	17,550	19,860	21,870	23,880	27,000	32,440	46,080
1940	24,000	21,300	36,600	41,630	46,380	50,550	57,010	66,950	80,700
41	4,320	3,780	6,850	8,860	10,150	11,480	14,420	16,073	18,087
42	8,600	6,480	9,880	12,920	14,860	16,240	18,230	20,225	27,716
43	23,300	19,800	26,220	31,450	34,740	37,370	46,070	51,330	63,820
44	12,800	10,600	14,080	17,470	21,480	27,460	31,830	35,500	39,196
45	28,000	22,800	38,000	45,600	48,170	56,410	67,870	78,350	94,090
46	10,400	7600	12,100	14,750	17,760	20,410	23,170	26,900	31,901
47	11,800	9,340	12,800	16,240	18,720	19,740	22,770	27,020	40,480

Note: Recorder installed on Sept. 20, 1929

9-230

UNITED STATES
DEPARTMENT OF THE INTERIOR
GEOLOGICAL SURVEY

Flood volumes of Licking River at Toboso, Ohio–arranged in order of magnitude.

Order of magnitd	Recurr. interval	Peak dischrge	Total runoff on consecutive days: 1 day	2	3	4	5	7	10	15
1	27	28,000	22,800	38,000	45,600	60,120	70,120	83,880	102,280	135,570
2	13.5	25,000	21,700	36,600	41,800	48,170	56,410	67,870	78,350	94,090
3	9	24,000	21,300	31,620	41,630	46,380	50,550	57,010	66,950	80,700
4	6.75	23,600	19,800	30,500	34,960	37,840	42,200	46,070	53,530	63,820
5	5.4	23,300	15,700	26,220	31,450	34,740	37,370	45,640	51,330	58,968
6	4.5	20,000	15,400	23,770	28,290	32,310	37,190	41,850	48,630	58,490
7	3.86	20,000	13,700	23,140	26,600	31,780	35,690	41,600	46,640	57,450
8	3.38	16,800	13,600	20,400	26,540	31,080	34,520	38,550	45,660	57,190
9	3.0	16,300	13,400	19,800	25,780	30,650	33,610	37,690	45,610	52,245
10	2.7	15,000	13,400	19,600	25,250	28,550	32,210	37,380	41,500	51,650
11	2.46	13,100	11,700	18,940	24,690	28,510	31,390	37,300	41,295	51,208
12	2.25	13,000	11,600	17,890	24,420	27,270	29,060	35,970	40,850	49,110
13	2.08	12,800	10,600	17,550	23,680	26,680	28,520	31,830	38,820	46,080
14	1.93	12,400	10,500	16,630	20,450	23,430	27,460	31,440	35,500	43,957
15	1.80	11,800	9,340	15,400	19,860	22,970	25,970	30,920	34,875	40,480
16	1.68	10,400	9,020	14,080	17,730	21,870	25,300	29,350	34,394	40,175
17	1.58	10,100	8,180	13,730	17,470	21,480	23,880	27,000	32,440	39,196
18	1.50	9,740	8,030	12,800	16,700	18,380	20,410	23,170	27,020	38,260
19	1.42	9,740	7,750	12,100	16,240	18,270	19,740	22,770	26,900	38,127
20	1.35	9,380	7,600	11,960	14,750	17,760	19,560	21,520	25,650	31,901
21	1.29	8,600	6,480	9,880	12,920	14,860	16,240	18,230	20,225	27,716
22	1.23	6,090	4,780	8,390	9,930	11,160	12,082	14,240	16,073	24,178
23	1.17	4,720	4,720	7,210	9,270	10,980	11,748	13,810	15,958	18,087
24	1.13	4,660	4,660	7,170	8,930	10,150	11,480	13,079	14,196	16,584
25	1.08	4,460	3780	6,850	8,860	10,130	11,330	12,490	13,220	14,599
26	1.04	4,320	3600	5,030	5,848	6,656	7,222	8,000	9,003	10,263

FIGURE 2.—Tabulation of flood-volume data.

secutive) discharge, 3-day, 4-day, 5-day, 7-day, 10-day, 15-day, or other period, discharge in each water year. A tabulation of flood volumes for Licking River at Toboso, Ohio, is presented as figure 2.

HISTORICAL DATA

Historical floods provide probably the most effective data available on which to base flood-frequency determinations, and where the data are reliable this information should be given the greatest weight in constructing the flood-frequency graph. Effort should be expended to search out historical data from newspapers, local historical society records, local history reports and other sources.

Historical data are particularly valuable where there is an account of all floods, above a certain stage, over a long period antedating the beginning of stream gaging. The minimum or base stage is the stage where damage begins or threatens. A list of historical floods is of the nature of a partial-duration series above a high base, but because there is generally only 1 flood of such magnitude in any 1 year, it may also be viewed as a partial list of annual floods. Where there is only one historical flood, the "maximum known," the base is the same as the flood.

An example of historical floods above a base is given by the record for the Susquehanna River at Harrisburg, Pa., where floods above 18.0 feet date from 1786 (Benson, 1950). If combined with the continuous record obtained in recent years, a period of record for 172 years is available.

A good example of the use made of newspaper sources for historical information is given by Woolley (1946). Most flood reports published by the Geological Survey contain data on historical floods; the report by Grover (1937) has information dating back to 1639.

Much historical information has been obtained in some areas, and incorporated in the annual reports on streamflow. (See reports on "Surface Water Supply of the United States, part 8, Western Gulf of Mexico Basins.") Examples are: (a) San Antonio River at San Antonio, Tex. (gaging-station record began 1915), "Maximum stage known since at least 1819, that of Sept. 10, 1921; flood of July 5, 1819, equalled or exceeded that of Sept. 10, 1921."; and (b) Nueces River below Uvalde, Tex. (gaging-station record began 1939), "Maximum stage known since at least 1836, 40.4 ft June 14, 1935, from floodmarks (discharge, 616,000 cfs, by slope-area measurement at former site.)" The value of such information in a flood-frequency study is obvious.

PLOTTING POSITIONS

The analysis of a series of flood data starts with a listing of the peaks under consideration; for example, all the annual peaks at a gaging station. These are ranked according to magnitude, customarily starting with the highest as 1 (col. 5, fig. 1). Some measure of frequency must then be computed so that a "plotting position" is obtained for the frequency scale. This plotting position is in terms of years. The objective of the frequency analysis is to determine the magnitude of the flood which will be equaled or exceeded once in a specified period of years; this specified period of years is known as the recurrence interval.

Based on considerations of probability, several methods have been used to compute the recurrence interval.

METHODS

"California" method.—This is the simplest form of computation, in which the recurrence interval, $T_r = \frac{n}{m}$, where n is the number of years of record and m is the rank starting with the highest as 1. (See Calif. Dept. Public Works, 1923; Jarvis and others, 1936). A minor objection to this method, where considering the probability of the event, is the reciprocal of the recurrence interval. The probability of the lowest flood occurring is computed as 1, which precludes the occurrence of any flood lower than this (such a point could not be plotted on probability paper). Also, this method gives no weight to the probability that the highest flood of record has a recurrence interval of something over n years.

Hazen method.—Hazen (1930) computed the "return period," T_p, as $\frac{2n}{(2m\text{-}1)}$. This equation results in a recurrence interval of $2n$ for the highest flood of record, which is an artificial lengthening of the period of record.

Gumbel method.—The theoretical plotting by Gumbel (1945a) is based on the assumption that the observed mth value is the most probable, or modal, value of this rank of flood. Its return period is therefore skewed towards the mode of the theoretical distribution.

The Gumbel theory does not apply strictly to floods for the following reasons:

1. It is assumed that the same treatment derived for daily discharges can be applied to peaks.
2. The daily discharges are not independent events.
3. The 365 daily discharges in a year do not constitute a "large" number as predicated by the theory.
4. The annual peaks under consideration do not come from the same statistical population. Some peaks are caused by ordinary seasonal rains, some by snowmelt, others by hurricane conditions. Entirely different physical factors influence each type. There-

fore, although other considerations fitted the assumptions, the different peaks at one station would not lie on a straight line.

In spite of this, the Gumbel plotting has some value because in certain ranges, particularly the low range, the frequency curve tends towards a straight line, so that an index flood chosen in that range is better defined.

Beard method.—This method (Beard, 1952) is based on the assumption that the *m*th value is the median value of this rank of flood, or that the probability of this value occurring is 50 percent.

Geological Survey method.—The formula used by the Geological Survey is:

$$T=\frac{n+1}{m}$$

where

T=recurrence interval, in years
n=number of years of record, and
m=magnitude of flood, the highest being 1.

This formula gives essentially the same results as Gumbel's computed values, and is much simpler to use. It has been adopted by Gumbel and by many adherents of his theories. (Gumbel, 1945b; Chow, 1953; Velz, 1952.) An explanation of the derivation of the formula is given by Langbein. (See p. 48.)

Recurrence intervals computed by this formula give results similar to those computed by the California method, but lack the theoretical deficiencies of the latter. For example, consider a record to contain 20 floods, numbered from 1 for the highest to 20 for the lowest. The California method would consider that the highest flood would be equaled or exceeded $1/20$, or 5 percent, of the time, a logical procedure. But suppose the numbering is reversed, then all floods will be equal to or less than the highest $20/20$ or 100 percent of the time. Obviously the sum of the percentages that a flood can occur, be exceeded, or fallen short of, must be equal to 100. The Survey method shows the highest flood equaled or exceeded, $1/21$ or 4.76 percent of the time, and that all floods are equal to or less than the highest $20/21$ or 95.24 percent of the time. Thus it shows that a flood can occur, be exceeded, or fallen short of 100 percent of the time (Cross, 1946, p. 17).

The formula adopted by the Survey is simple to compute, is applicable both to annual flood data and to the partial-duration series, and gives results that acceptably conform with some of the latest theories.

PLOTTING HISTORICAL DATA

In computing plotting positions by any formula there are times when the computations must be modified. For example, the highest flood of a 40-year record should not always have a recurrence interval of 41 years. There is frequently additional historical knowledge which might indicate, for example, that the highest flood in a 40-year record is the highest in 300 years. Its plotting position should then be computed as though it were the highest in a series of 300 items or as 301 instead of 41 years. The second flood (in 40 years) would then be computed as usual.

Plotting positions for historical floods may be computed for three cases:

1. Where there is one historical flood, higher than any for the period of record, with no intimation of what may have occurred between the date of the historical flood and the beginning of the record except that the historical flood is higher than any other flood known. In this case, the recurrence interval of the historical flood should be recorded as equal to one plus the period for which it is the greatest. An example of this is the 1876 flood, listed in figure 1.
2. Where there is one historical flood known to be the highest until a greater flood occurs during the period of record. As an example, assume a discharge of 1,000 cfs occurred in 1863 and that the record began in 1923; the 1863 flood was the "maximum known" until 1938 when a discharge of 2,000 cfs was recorded.

Plotting positions, up to and including the 1957 flood

Floods	Formula	Years	Cubic feet per second
Maximum flood in 1863–1957	$\frac{95+1}{1}$	96	2, 000
2d highest in 1863–1957	$\frac{95+1}{2}$	48	1, 000
2d highest in 1923–57	$\frac{35+1}{2}$	18	800
3d highest in 1923–57	$\frac{35+1}{3}$	12	600
35th highest in 1923–57	$\frac{35+1}{35}$	1. 03	100

3. Where there is a historical record of all floods above a high base, as above "bankful" stage, and it may be assumed that lesser floods follow the same distribution as for period of record. Order numbers, and recurrence intervals, for annual floods for the period of record can be adjusted to the longer period for which historical data are available by a method described by Benson (1950). The general formula for the transformed order numbers is:

$$m_1=A+\frac{H-A}{T-A}(m-A)$$

where

m=order number, where the highest flood is 1, the second highest 2, and, for all floods both those for period of record and those from the historical record;

m_1=order number of floods below base of historical record, adjusted to the time base of the historical record;

A=the number of annual floods equalling or exceeding the lowest historical flood;

H=the length of the historical record, in years; and

T=the total number of items, historical and recent, in the array.

To illustrate, assume the same record as shown for example 2. There would be a total of 36 items (1863, 1923–57). Of these, the first 2 are above the base of 1,000 cfs and are the 2 highest in 95 years of record (1863–1957). The other 93 years of record since 1863, excluding these 2, are below 1,000 cfs. The other 34 items of known record below 1,000 cfs are assumed to represent the distribution of frequencies during the 93 years below 1,000 cfs. Each of the 34 years are considered equal to 93/34 or 2.735 years of record. The transformed orders of all items after the first two are therefore computed by the formula $m_1=2+93/34(m-2)$. The order of the last item is 95, and the total array now represents the frequency distribution in 95 years of record. The recurrence intervals for all 36 items are computed by using the formula $\frac{(95+1)}{m_1}$.

Plotting positions

Year	Discharge (cfs)	Order			T (years)
		m	m-2	m_1	
1938	2, 000	1	--------	1	96. 0
1863	1, 000	2	--------	2	48. 0
	800	3	1	4. 74	20. 3
	600	4	2	7. 47	12. 9
	100	36	34	95. 0	1. 01

An example for a record beginning in 1786 and continuous since 1874 (Susquehanna River at Harrisburg, Pa.) is given by Benson (1950).

PLOTTING PAPER

For the general purposes of flood-frequency graphs, the kind of graduations on the paper is not important. However uniformity is desirable, and if a choice is to be made, the chart based on the theory of largest values has much to offer. A time scale has been devised by Powell (1943) that tends to make the frequency curve plot as a straight line. The equation for the time graduation is:

$$y=-\log_e\left[-\log_e\left(1-\frac{1}{T}\right)\right]$$

where
y=a linear distance
T=recurrence interval.

The zero value of y occurs at T=1.582, and other values must be measured from this zero point.

Values of T *and* y

(Powell, 1943)

T	y	T	y
1. 01	−1. 53	5. 0	1. 50
1. 05	−1. 10	10	2. 25
1. 25	−. 48	20	2. 97
1. 67	. 09	25	3. 20
2. 00	. 37	50	3. 90
2. 33	. 58	100	4. 60
2. 50	. 67	200	5. 30
3. 33	1. 03	500	6. 21
4. 00	1. 25	1, 000	6. 91

These values of T and y may be used to prepare plotting paper; generally recurrence intervals are measured along the horizontal axis. The y values may be measured in inches, or in inches times a constant to adjust to different sizes of paper.

Two discharge scales are used for plotting annual floods, one rectangular and the other logarithmic. Standard Survey forms have been printed for each of these two scales; these are not available generally. Survey form 9–179a is to rectangular ordinate scale and form 9–179b is to logarithmic ordinate scale; these are designed so that the time scale may be extended by attaching standard 3-inch logarithmic graph paper.

Annual flood peaks are plotted to either rectangular or logarithmic discharge scales, whichever provides the nearest to a straight-line plot. (A sample of these forms, giving examples of data plotted to both scales, is shown on figures 3 and 4.) See pages 22, 23.

For partial-duration series plots, 3-cycle semilog graph paper generally is used. The linear scale (ordinate) is used for the discharge

540570 O - 60 - 4

data and the logarithmic scale (abscissa) for recurrence interval. (A duration-series form giving an example of a data plot, is shown on figure 5.) See page 24.

FITTING FREQUENCY GRAPHS

When a frequency diagram is plotted there is a need for fitting a curve to the data. The fact that most streamflow records are less than 25 years in length does not, however, satisfy the demand for estimates of long-term destructive floods. The tendency is to use the frequency graph for purposes of extrapolation; this tendency cannot be encouraged. The linear distance from 25 to 200 years seems very short on most graphs, but extrapolation can only be justified when the phenomena have been proven to conform to underlying law. The error of a curve fitted by whatever method may be extremely great at its outer end. Since no known fitted curve can serve any use in extrapolation, its main purpose would be merely to provide a smoothing or interpolation formula. The value of an analytically fitted function therefore seems doubtful.

Many distrust any method that allows for personal judgment and prefer a mathematical fitting of curves. In order to use mathematics, it is first necessary to know the mathematical law to which flood data must conform—such a law has never been demonstrated. The many studies of flood-frequency relations made so far have failed to define a specific distribution that is typical of all locations.

Again, the use of a least-squares method of fitting a straight line to flood data may lead to absurd results, because of the neglect of hydrologic factors. A good example of this is the short period of record containing a high flood—the single high flood may outweigh all other floods and lead to a straight line which is away from any of the plotted points. In such a case, although the lines drawn graphically by a number of men might vary, each would be better than a mathematically fitted line. A method is not better because it leads to uniform answers, if those answers are uniformly wrong.

Only graphical treatment is contemplated for Survey reports. The several plotted points should be indicated by circles.

Unless a very long record is being analysed, as at major stations on large rivers, it is not advisable, except in an emergency, to prepare a frequency curve derived from a single station. The array of peaks at any one station is a random sample, and, as such, may be far different in character, particularly in a short record, from those of nearby stations of otherwise like characteristics. Flood frequencies should be generalized and related to a common period of record where possible. Frequency characteristics of individual stations should then be based on or related to these generalized frequency curves.

For certain engineering applications, where single station analysis only is done, a simple curve of best fit may be drawn. The recurrence interval for the highest flood is doubtful, and those next highest probably decreasingly so, hence the topmost points cannot always be used directly in drawing the curve. The design of the plotting paper tends to straighten out the shape of the frequency curves, but, with the above exception, the base data remains the best indication of the shape.

If it becomes necessary to extrapolate any appreciable distance beyond the extent of the usable data, then methods other than mere extension of a curve become necessary. These methods might include study of the ratios of the 25- or 50-year flood to the mean annual flood at several long-term stations. They might include a compilation and graphing of extreme discharges in an area, plotting discharge against drainage area. From an envelope curve for such data, to which some frequency might be assigned, based on composite length of experience or frequencies attributed to some of the data, a figure might be obtained which could be used either as a point towards which to extend the curve or as a value which it might approach asymptotically.

Examples of frequency curves based on the annual flood series, plotted to both rectangular and logarithmic scales, and the partial-duration series, plotted to semilogarithmic scales, are presented as figures 3, 4, and 5, respectively.

An example of volume-frequency curves, plotted from data of figure 2 is presented as figure 6.

SINGLE-STATION ANALYSIS

The foregoing discussion on methods of frequency analysis applies to the analysis of the records at an individual site. The results of such a study represent an exact description of what has happened at the site in the past, for some definite period of time.

However, the purpose of frequency analysis is generally the prediction of what will happen in the future. In using past records to predict the future, it must be assumed that there is no change in the nature of the factors causing floods. The past record is then considered a sample of the total statistical population consisting of past and future floods. However, when used for predicting the future, the sample must be considered as only an approximation, because all samples may vary from the group as a whole. The extent of the variation, that is, the probability of any extent of variation from the norm, is known from statistical theory.

In general, flood records represent relatively short samples, statistically speaking. For this reason, the record at any individual station may depart considerably from a true representation of the overall long-time flood-frequency relation.

A study was made to demonstrate the variability inherent in short records by Benson (1952). Basically the same report is reproduced on pages 51–74. This study was made starting with a theoretical

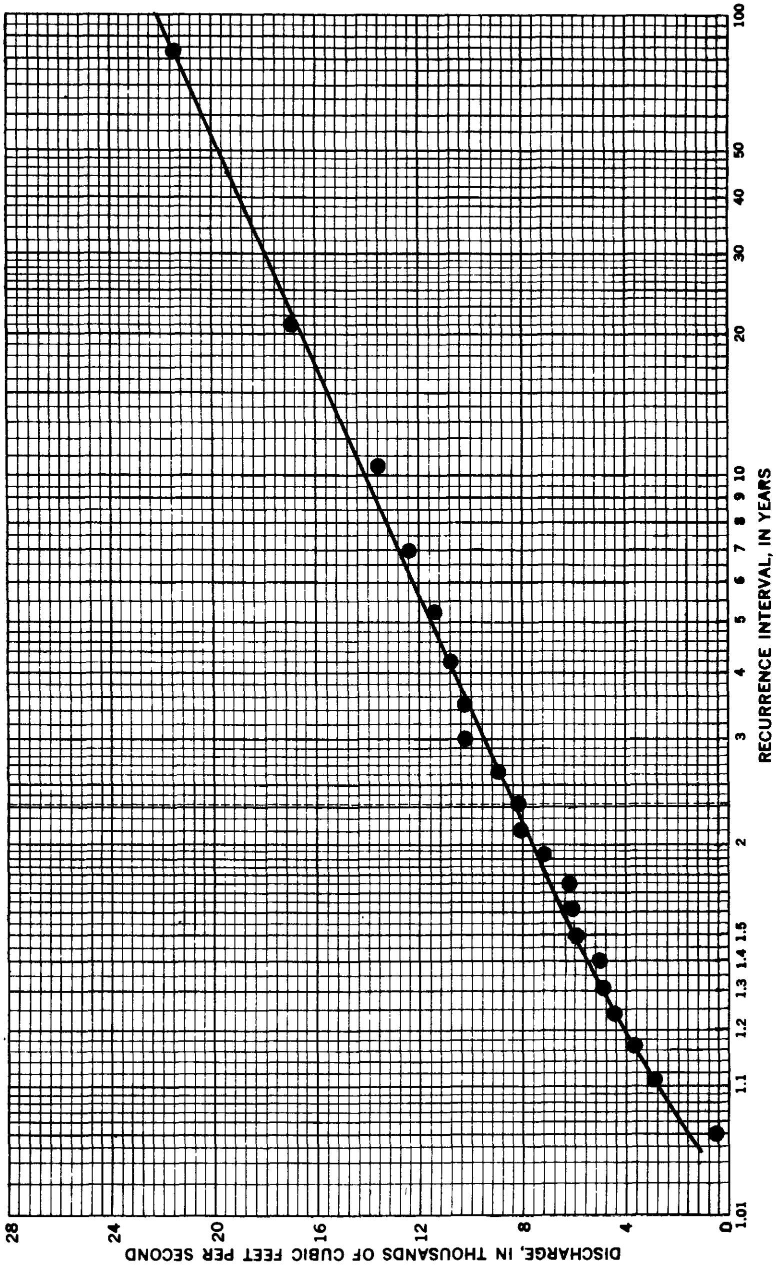

FIGURE 3.—Flood-frequency curve, rectangular scale.

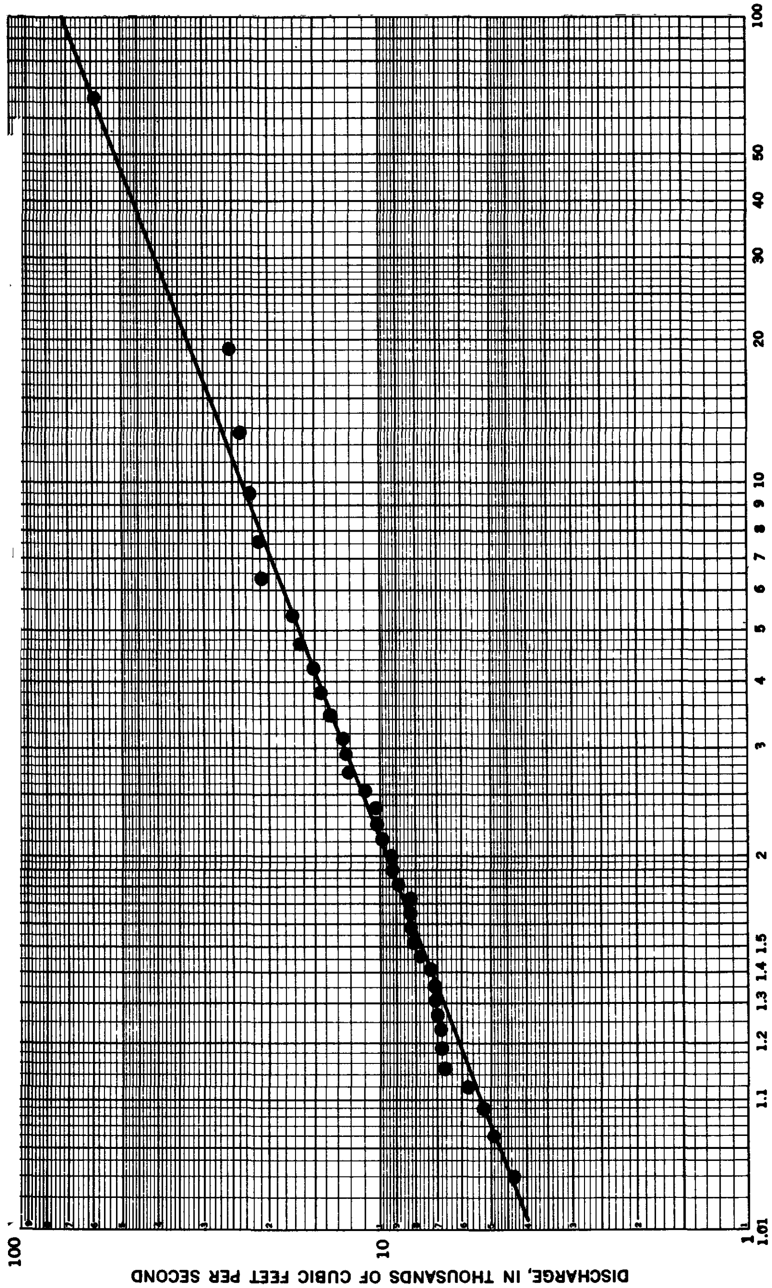

FIGURE 4.—Flood-frequency curve, logarithmic scale.

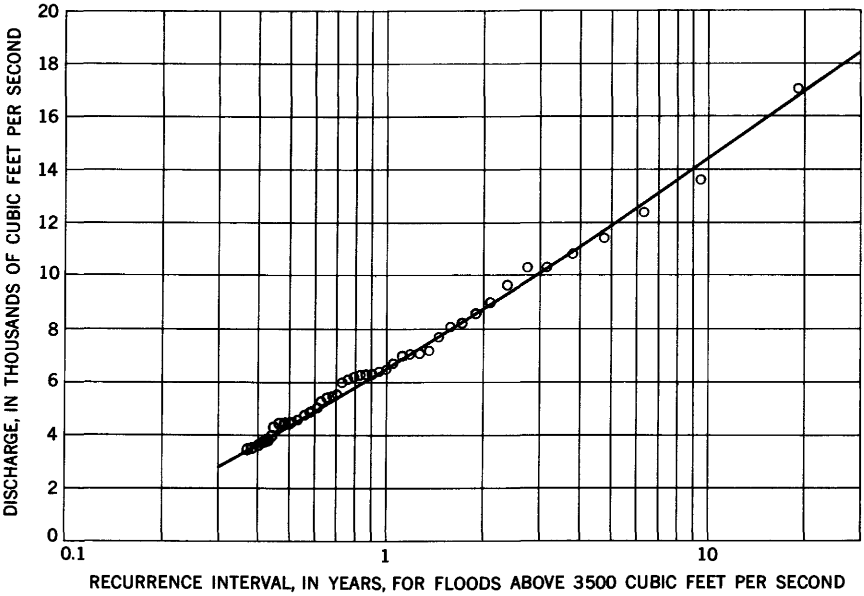

FIGURE 5.—Partial-duration series flood-frequency curve.

1,000 years of record with annual peaks distributed so as to exactly define a straight-line graph on the standard frequency graph paper. Individual peaks were then drawn at random to form groups of samples equivalent to 10-year, 25-year, 50-year, and 100-year periods of record. Frequency curves for these short-term periods give a graphic picture of the variation in frequency curves for short-term samples (figs. 20–23). Even the 100-year records show more variation than might be expected. The significance is that an actual gaging-station record of 25 years, for example, could take the position of any one of the curves in figure 21. The range in variation from the long-term value at any frequency can be judged from these figures. Even the mean annual flood (recurrence interval, 2.33 years) can vary significantly from the overall value. The study shows that 12 years of record are required to define the mean annual flood within 25 percent (with expectation of correct results 95 percent of the time).

Because of the unreliability of the single-station curve, more extensive methods have been developed by the Survey. These include a study of frequency relations on a regional basis. The results of such studies reduce the large sampling errors in individual records and are applicable to all points within the region studied.

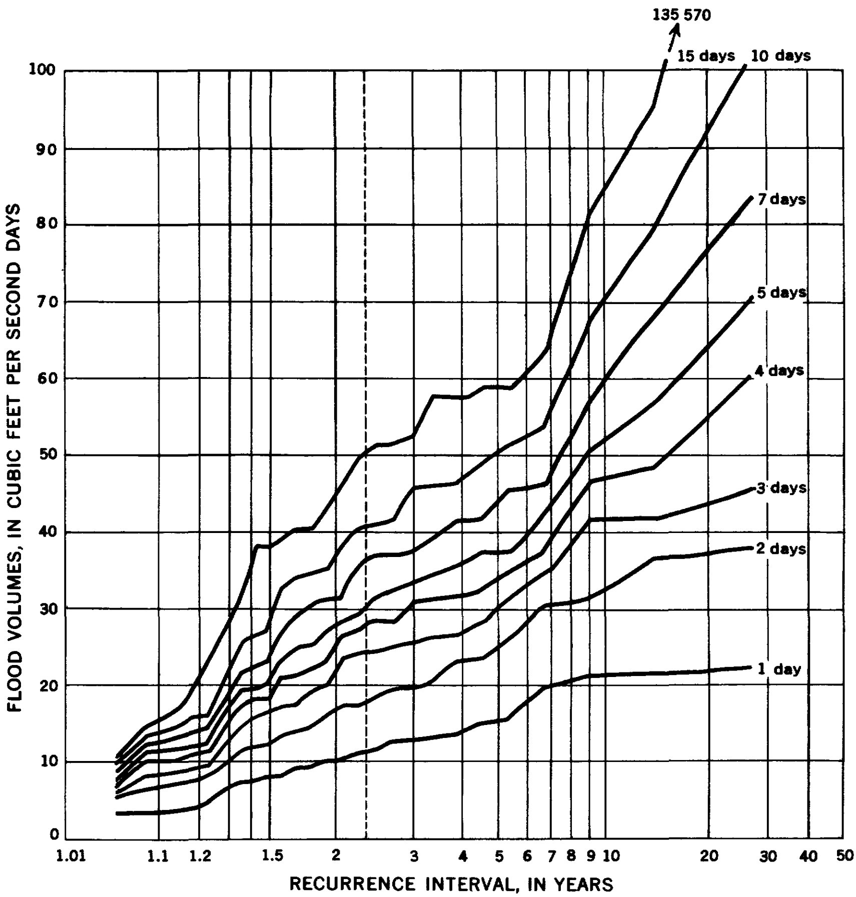

FIGURE 6.—Annual flood volume-frequency curves.

REGIONAL FLOOD FREQUENCY

Flood records at gaging stations are generally short, the sampling errors correspondingly large, and the records represent different periods of time. In addition it is only rarely that flood-frequency information is needed at a gaging-station site. More often it is required at an ungaged site.

A method of combining records within a region has been developed which reduces the sampling error, bases the results on a uniform period of experience, and produces flood-frequency relations generally applicable within the region. Experience has proved that such a procedure is not only possible but leads to results of acceptable accuracy for design and planning.

It has been suggested that combining records by the station-year method of analysis could be applied profitably to flood-frequency studies as it has been to rainfall-intensity frequencies. By this method, for example, 5 records of 20 years each could be combined to obtain a 100-year record. Such a method requires that flood-frequency characteristics be comparable and that the data be independent. Studies indicate that the first of these requirements is met by some drainage basins or regions but the second requirement is not met, at least by gaging stations with drainage areas of over 100 square miles. Five stations in a hydrologically homogeneous area measure the same flood 5 times rather than measuring 5 floods each year; in other words, a 20-year period cannot yield more than a 20-year record, and cannot be expanded so as to become a 100-year record.

Rather than add several short records to produce a long-term record, they may be averaged thereby providing several measurements of each year's event. By this method, 5 records of 20 years each when combined give only a 20-year record, but it is considered that each year's flood has been measured 5 times. The median of these five values is assumed to give a better measure of the frequency characteristics of that event.

In order to test this assumption, a study was made of a theoretical 1,000-year record (Benson, 1952). A perfect 1,000-year flood history was assumed, and divided into 100 10-year periods, each randomly distributed. One hundred frequency curves were drawn, each based on a 10-year record. The variation of these 100 curves from the basic 1,000-year curve gives a measure of possible variation of curves based on short-term records of natural events. These 100 10-year frequency curves scattered greatly from the true curve.

By combining only 5 or 10 of the 10-year records, and selecting the median, it was found that the resultant composite curve reduced the chance for error greatly. The study indicated that by combining 5 records the chances are better than 50–50 that the composite curve will lie within the central one-fifth of the spread of all the individual curves and by combining 10 records, chances are 3 out of 4 that it will do so.

Regional flood-frequency study consists of two major parts. The first is the development of basic, dimensionless frequency curves representing the ratio of the flood of any frequency to an index flood (the mean annual flood). The second part is the development of relations between topographic characteristics of drainage areas and the mean annual flood, to enable the mean annual flood to be predicted at any point within the region. The combination of the mean annual flood with the basic-frequency curve, which is in terms of the mean annual flood, provides a frequency curve for any location.

BASIC-FREQUENCY CURVE

Throughout large regions, which are homogeneous with respect to flood-producing characteristics, individual streams having a wide range of drainage area will have frequency curves of about equal slope or steepness. If the peak floods at each gaging station are reduced to dimensionless ratios by dividing by an index flood, the curves plotted using the flood ratios can be superimposed and will nearly coincide.

These curves will all pass through the 2.33-year frequency at the ratio of 1.0, but will have somewhat different slopes. The variation of these slopes can be tested to see whether the spread could occur by chance among samples from the same population (this is the "homogeneity test," which uses the ratio of the 10-year flood to the 2.33-year flood as the slope). If so, the assumption of a homogeneous region containing all the stations is satisfactory. Within such a homogeneous region, the best representation of the flood-frequency relation can be obtained by combining all dimensionless curves. The resulting average-frequency curve is then applicable throughout the region, and is called the regional frequency curve.

If the spread in slope is greater than can be attributed to chance alone, then further separation of the region is made to produce two or more homogeneous regions. Statewide studies have commonly found 2 to 3 such regions within a State; however, these will correlate with like regions in adjacent States, so that the area of a particular homogeneous region can be large.

Combining individual curves to obtain the regional frequency curve is accomplished by taking the median of station-flood ratios at the same recurrence intervals. These median values are then plotted to define the regional frequency curve.

HOMOGENEITY TEST

The homogeneity test was developed by Langbein of the Geological Survey and is explained below.

The question whether the records in a group are homogeneous may be answered in a statistical sense by determining whether they differ from one another by amounts that cannot reasonably be expected by chance. Naturally no two records precisely represent the same experience nor have exactly comparable characteristics. On the other hand, where these differences are no more than those due to the operations of chance, we can readily conclude that they merely represent different aspects of the same thing and thereby group them.

A chart such as is shown for the homogeneity test could be constructed as follows: If a 1,000-year record were divided into 100 10-

year records in which the highest flood in each of the 10-year periods represented an estimate of a 10-year flood, the distribution of those 10-year floods could then be studied. They would not all be the same but the differences that exist would be merely due to chance. From this record, it could be decided what ranges of variations could be accepted and what variations would represent results other than chance. The same test could be done by studying a 2,000-year record in which 100 20-year records could be charted. In each chart the second highest flood would represent the 10-year flood. The spread of the 10-year flood as estimated for 20-year records would be less than in the second chart. From this the kind of spread due to chance could be determined.

There are no records of such length, but such a record is not necessary because if it is agreed that floods can occur fortuitously, then it is possible to calculate from theory the distribution of floods in a 1,000-year or 10,000-year record. This cannot be done in terms of discharge but it can be done in terms of probability. Such calculations are the basis of the test for homogeneity.

The "true" position of a frequency graph for a station can differ greatly from the position indicated by the plotted points. Although the true position is unknown, the chances that it will lie within certain distances from the plotted position can be calculated. Therefore, if several frequency graphs are plotted, most of them will group together, but there will be certain ones that plot apart. The chance that these variants represent can be calculated. If the calculations show that the spread is too great, those graphs are set aside as not being a part of the group being tested. The problem is to calculate the spread to be expected, and to set a limit to the spread that will be acceptable.

The standard deviation of the reduced variate, y, of the Gumbel distribution is equal to:

$$\sigma y = \frac{e^y}{\sqrt{n}}\sqrt{\frac{1}{T-1}}$$

where

T=recurrence interval
n=number of years of record
y=function of the recurrence interval T

This means that in a large number of different but homogeneous records, each n years long, probably two-thirds of the estimates of the T-year flood will be within σy of their most probable value of T.

A wider range of variation has been decided on and therefore two standard deviations have been selected as the permissible range. This means that 95 percent of the estimates will lie within 2 σ of the most probable value of T. It was also decided to make the test on the 10-year flood because this is the longest recurrence interval for which most records will give dependable estimates. The following calculations are derived ($T=10$ years; $y=2.25$; and $e^y=9.49$.):

$$2\,\sigma y=\frac{2e^y}{\sqrt{n}}\sqrt{\frac{1}{T-1}}\cdot\left[\text{or } 2\,\sigma y=2\sqrt{\frac{T}{n}}, \text{ very closely}\right]$$

$$2e^y\sqrt{\frac{1}{T-1}}=2\times 9.49\sqrt{\frac{1}{9}}=6.33.\ [\text{or } 2\,\sigma y=2\sqrt{10}=6.32]$$

A table giving values of y corresponding to T can be found in Gumbel (1942; also Powell, 1943).

Values of y *corresponding to* T

[Columns headed T_L and T_U respectively give the lower and upper limits of the chart for the homogeneity test]

n (in years)	y	$\frac{6.33}{\sqrt{n}}=2\sigma$	Lower limit		Upper limit	
			$y-2\sigma$	T_L	$y+2\sigma$	T_U
5	2.25	2.84	−0.59	1.2	5.09	160
10	2.25	2.00	+.25	1.85	4.25	70
20	2.25	1.42	.83	2.8	3.67	40
50	2.25	.90	1.35	4.4	3.15	24
100	2.25	.63	1.62	5.6	2.88	18
200	2.25	.45	1.80	6.5	2.70	15.5
500	2.25	.28	1.97	7.7	2.53	13
1000	2.25	.20	2.05	8.3	2.45	12

MEAN ANNUAL FLOOD

The mean annual flood for a gaging station is by definition, the 2.33-year flood from the graphic-frequency curve defined by points which are referred to the same base-time period.

The magnitude of the mean annual flood may be affected by many factors, which can be classed as either physiographic or meteorologic. The problem is: Given a drainage basin of certain physical characteristics, located in a region subject to certain meteorologic conditions, what mean annual flood can be expected? The answer is obtained by correlating the known mean annual floods of drainage areas within a region with the known characteristics of the basin and the region.

PHYSIOGRAPHIC FACTORS

The physiographic factors which may influence the mean annual flood at a given point are: (a) The size of drainage area, (b) channel storage (c) artificial or natural storage in lakes or ponds, (d) slope of streams, (e) land slope, (f) stream density, (g) stream pattern, (h) elevation, (i) aspect, (j) orographic position, (k) underlying geology, (l) soil cover, (m) cultivation, and others.

Practical methods of computing some of the important physical factors are described by Langbein and others, (1947). Some of the factors listed are fairly simple and can be expressed by a definite figure. Others, such as geology, channel storage, or orographic pattern are difficult to evaluate and have not yet been successfully used in correlations with peak floods. Many of these factors are interdependent.

Development of the relations, if they exist, requires much work. It also requires much work to compute them for ungaged areas. In many cases, topographic maps are not available for computing topographic characteristics.

METEOROLOGIC FACTORS

Meteorologic factors are concerned with the magnitude and distribution pattern of the precipitation falling on a drainage area. Some of the elements involved are: (a) Type of region, whether humid or arid, (b) storm directions, (c) storm patterns, (d) storm volumes, (e) precipitation intensities, (f) effect of snowmelt, (g) extent of ice jams, and probably others.

The evaluation, treatment, and use of the meteorologic elements are generally less certain than for the physiographic factors. The difficulty lies in determining what precipitation figures to use. Total annual precipitation has been used, but this is related only generally to storm rainfall. Rainfall intensities would be more directly related to peak discharges, but intensities must be expressed by both a definite period of time and a frequency of occurrence, as for example, a 50-year, 5-minute intensity. The possible combinations are many, and since this is only one of many other factors, both topographic and meteorologic, the selection of the best parameter becomes difficult. There is a great deal of opportunity for original investigation in this field.

COMPOSITE FACTORS

Many physiographic and meteorologic factors make demonstrating significant correlation difficult except for those factors that are outstandingly influential.

Various combinations of previously listed factors have been used in correlation with mean annual floods. One such combination is the mean annual runoff. This is a general index of the amount of precipitation available, and also an indication of the runoff-inducing characteristics of the basin. The mean annual rainfall is another factor which has been used, although not as successfully as the runoff.

In the Illinois flood-frequency study by Mitchell (1954) the basin lag was used. This is the time lag between the center of rainfall and the center of runoff. This time lag represents the composite effect of most or all of the topographic factors. It is therefore a very useful figure, but, it cannot always be defined from known basin characteristics. It must be measured during actual storms.

Another method equal to using a composite factor is that of dividing the study region into several parts, called hydrologic areas. Within each area a separate curve of mean annual flood is correlated with the drainage area and perhaps some other significant factor. In each of these areas such factors as rainfall, geology, and other features probably have the same overall effect.

If satisfactory relations for several areas can be found by using drainage area alone, it may not be worthwhile to include other factors which would improve the correlation only slightly. The data are generally limited, and the additional factors cause a loss of degrees of freedom, so that no improvement results.

If, after all practicable factors have been considered in the correlation, then the residuals from the average relation may be analyzed for geographic location. A pattern may result which can be associated with orographic effect, soils, or some other factor. If isograms of the residuals are plotted, the use of a mean coefficient may improve the accuracy of the mean annual flood.

COMPUTATION PROCEDURE

SELECTION OF STATIONS

The first step in beginning a flood compilation is to list the gaging stations. A record should be included if it is 5 or more years in length, although generally recurrence intervals should not be computed for records shorter than 10 years. Old, discontinued records that otherwise qualify should be included. Storage or other artificial factors which would tend to modify flood discharges significantly should be a minimum. Exclude canals, ditches, and drains, in which discharges are subject to substantial control by man. Always include the total usable storage capacity in the basin above the gage.

If the records for 2 drainage areas on the same stream show differences in area of less than 25 percent, the 2 may be treated as 1 record. If both records are for the same period of time, use the better one or use both and give each a weight of one-half; if the records are for different periods of time, combine them into one longer record. This is not always done for the larger rivers, such as in the lower reaches of the Missouri; see the treatment given to the Mississippi River (Searcy, 1955, p. 12) and a later discussion on page 46.

Include stations maintained by other agencies, such as the U.S. Weather Bureau; Corps of Engineers, Department of the Army; U.S. Soil Conservation Service; State agencies; private agencies, such as power companies, water companies, and milling companies; and any others. Include stations at which only stage records have been kept.

Use all records representing natural flow. Do not discard a record because it does not fit into a pattern defined by other records.

List the stations not included in the compilation because they have too short a record, are not rated, or for other reasons; show why these stations were excluded.

The step-by-step procedure for making a regional flood-frequency analysis will be illustrated by the study made in the Youghiogheny and Kiskiminetas River basins, in Pennsylvania and Maryland (Noecker, 1952).

In the 3,800 square miles drained by the Youghiogheny and Kiskiminetas Rivers, 15 gaging stations had annual flood records of 11 to 69 years when the analysis was made in 1951; these were all the records of adequate length that were not seriously affected by artificial regulation. These records were used in the analysis; the stations are:

1. Stony Creek at Ferndale, Pa.
2. Conemaugh River at Seward, Pa.
3. Conemaugh River at Tunnelton, Pa.
4. Kiskiminetas River at Avonmore, Pa.
5. Little Conemaugh River at East Conemaugh, Pa.
6. Blacklick Creek at Blacklick, Pa.
7. Loyalhanna Creek at Kingston, Pa.
8. Loyalhanna Creek at New Alexandria, Pa.
9. Youghiogheny River at Ohiopyle, Pa.
10. Youghiogheny River at Connelsville, Pa.
11. Youghiogheny River at Sutersville, Pa.
12. Casselman River at Markleton, Pa.
13. Big Piney Run near Salisbury, Pa.
14. Laurel Hill Creek at Ursina, Pa.
15. Green Lick Run at Green Lick Reservoir, Pa.

The locations of these stations are shown on figure 7.

Records for six other stations in the area were not used because of excessive backwater (making the discharge record doubtful or fragmentary) or because they were of too short duration (5 years or less).

SELECTION OF BASE PERIOD

Any statistical analysis of flood data must proceed on the assumption that the data analyzed are of a random nature. Yet there are persisting patterns or cyclic changes in weather, so the storms or floods of one period of time will not equal in general magnitude or distribution those of another period. Although the weather pattern may be random when considering a long period of time already past or in forecasting the future weather pattern, any definite time period in the past may be either high or low with respect to the average.

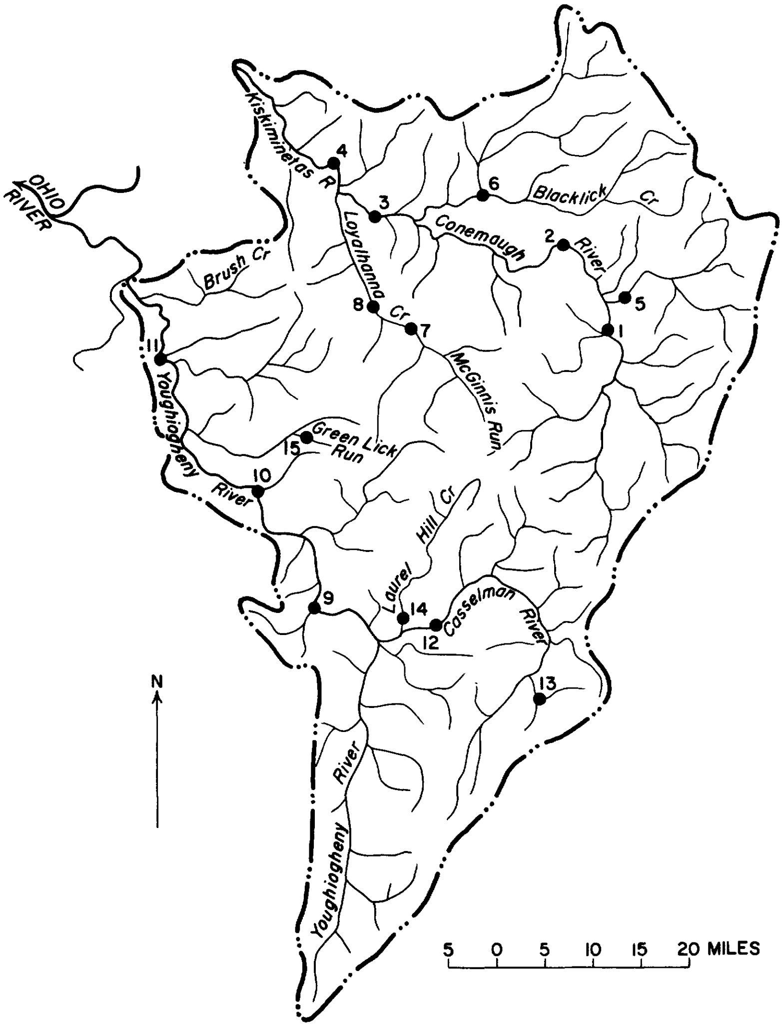

FIGURE 7.—Youghiogheny and Kiskiminetas River basins, showing location of gaging stations.

The relation of any particular period of time to the average is difficult or impossible to define. In order to combine records, comparable data must be analyzed. This eliminates the variability with time, so that the effect of other factors on flood peaks may be analyzed more easily. For the purpose of defining the long-term relation the period of time used should be as long as possible; for the purpose of separating the various factors affecting flood peaks, a common period of time for all records is desirable.

The base- or common-time period which is chosen is derived from the longest available records. Although all records will not be equally long, missing years may be filled in by correlation procedures. Such estimated peaks are not used directly, but serve the purpose of allowing the correct order numbers to be assigned to the peaks of record, with respect to the base time period. The records then can properly be compared or combined. A bar graph furnishes a compact record of what data are available, and is an aid in selecting a base period.

Bar graph.—A bar graph showing graphically the length of usable records in the example area is presented as figure 8.

The bar graph shows that, to avoid excessive adjustment of records, selection of two base periods would be desirable. One period was selected as 1914–50 and the other 1884–1950. The procedure was to extend all records to the short period, five records to the long period, and then to adjust from the short to the long period in one step. The procedure for making this adjustment is described on pages 40 and 42.

ADJUSTMENT OF RECORDS TO BASE PERIOD

The problem is to compute recurrence intervals for the floods in a partial record (not complete for the base period) that will be the same for these floods if the record was complete. The steps in the adjusting process follow.

List data on work sheets.—Annual flood peaks for each gaging station should be copied on work sheets, such as Survey form 9–179, that has a line for each year of the base period. Tabulate only the year and the discharge, as shown in columns 1 and 4 (for 1933–50) of figure 9. A similar form should be made for each gaging station in the region under study.

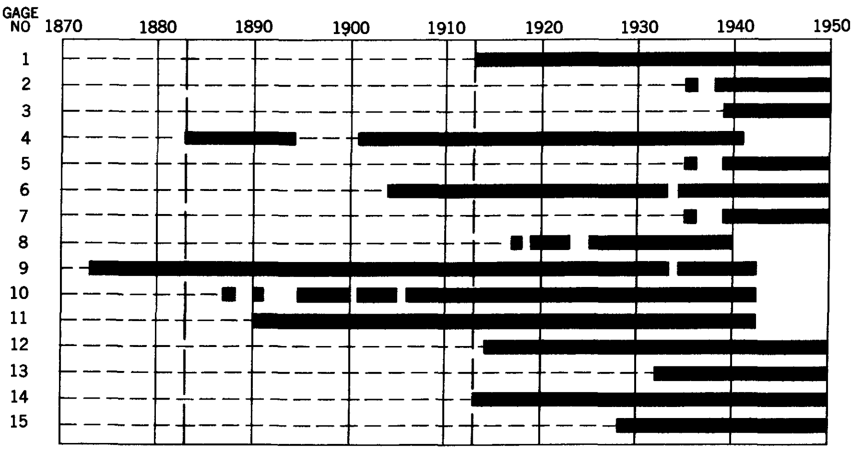

FIGURE 8.—Bar graph showing period of record of maximum annual peaks at gaging stations.

9-179
(October 1952)
FLOOD DATA

UNITED STATES DEPARTMENT OF THE INTERIOR
GEOLOGICAL SURVEY
WATER RESOURCES DIVISION

File

Flood data for (13) Big Piney Run near Salisbury, Pa.
Drainage area 24.5 square miles. Period ~~of record~~ 1914-50
Flood data for momentary peak discharges greater than — cfs.

WATER YEAR	DATE	GAGE HEIGHT (Feet)	DISCHARGE		ANNUAL FLOODS		PARTIAL DURATION SERIES		REMARKS
			CFS	Ratio to $Q_{2.33}$	Order (M)	Recurrence Interval (Years)	Order (M)	Recurrence Interval (Years)	
1	2	3	4	5	6	7	8	9	10
1914			(480)		35				
1915			(875)		22				
1916			(1,050)		16				
1917			(1,700)		7				
1918			(1,350)		11				
1919			(540)		33				
1920			(1,040)		17				
1921			(630)		29				
1922			(1,550)		8				
1923			(370)		37				
1924			(2,300)		4				
1925			(790)		25				
1926			(790)		26				
1927			(1,000)		18				
1928			(1,520)		9				
1929			(620)		31				
1930			(820)		23				
1931			(760)		27				
1932			(610)		32				
1933			1,900		5	7.60			
1934			968		19	2.00			
1935			494		34	1.12			
1936			4,100		2	19.0			
1937			4,300		1	38.0			
1938			1,860		6	6.33			
1939			1,180		14	2.71			
1940			1,510		10	3.80			
1941			1,260		13	2.92			
1942			1,290		12	3.17			
1943			3,460		3	12.7			
1944			628		30	1.27			
1945			1,050		15	2.53			
1946			908		20	1.90			
1947			634		28	1.36			
1948			876		21	1.81			
1949			472		36	1.06			
1950			796		24	1.58			

Graphical mean annual flood ($Q_{2.33}$) 1,060 cfs for period 1914-50

Sheet ____ of ____ Listed by T.D. Date 6/28/58 Checked by CHY Date 6-28-58

U. S. GOVERNMENT PRINTING OFFICE 16—57884-2

FIGURE 9.—Tabulation of flood data, Big Piney Run near Salisbury, Pa.

Estimate years of no record.—There is no record for the period 1914–32 at Big Piney Run. To complete this period, plot annual floods of record against annual floods at another station where the record is complete. Plot annual flood versus annual flood, not a flood at one station versus a flood of the same date at the other station, unless the annual floods occur at the same time. For example, figure 10 shows a correlation of annual peaks for Big Piney Run

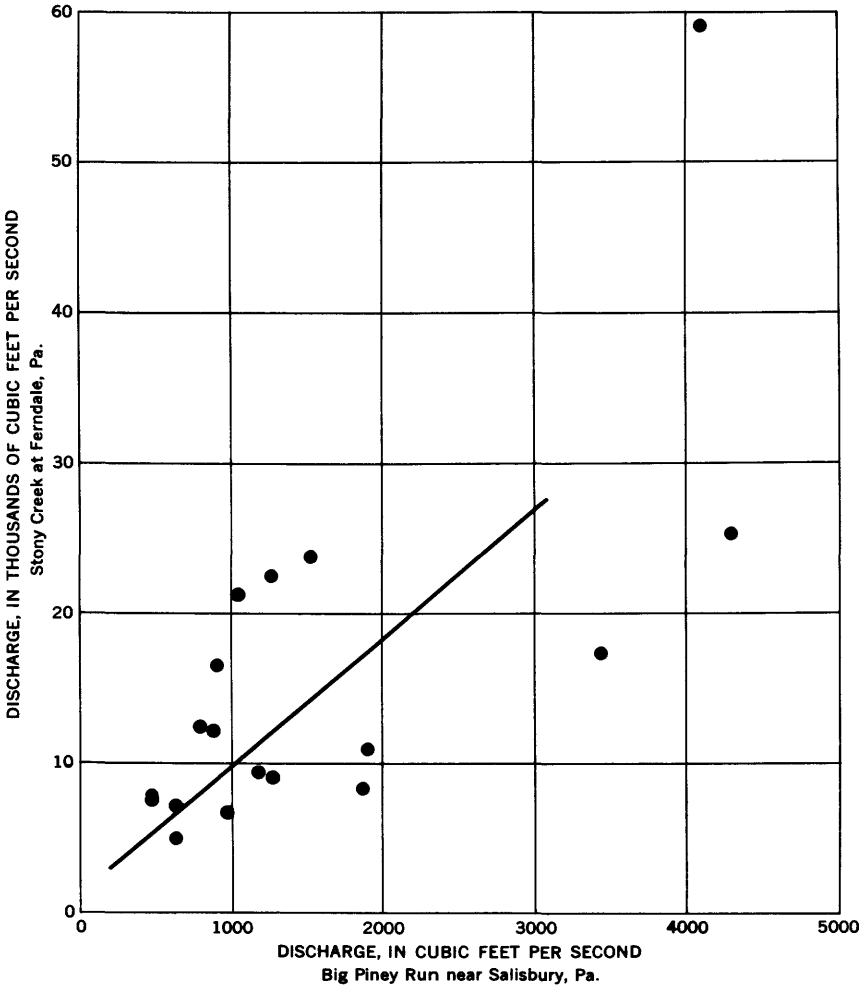

FIGURE 10.—Correlation of annual floods, Big Piney Run versus Stony Creek.

and Stony Creek. The correlation is not perfect; if it were, there would be no independence and there would be only one record that had been measured twice.

From the curve shown on figure 10, record the estimates for the years 1914–32 for Big Piney Run from the record at Stony Creek. These estimates are shown in parenthesis in column 4 of figure 9. Although discharge records have been used to obtain the estimates, do not plot or use them as discharges, but consider them as computation figures.

Compute order numbers.—There now is either an annual peak discharge or an estimate for each year of the base period. Order numbers should be assigned to all of these figures; the largest should be numbered 1, the second largest 2, and so forth. These order numbers are shown in column 6 of figure 9.

Compute recurrence intervals.—Next, compute recurrence intervals, from the formula $T=\frac{(n+1)}{m}$, for each flood of record. Do not compute recurrence intervals for the estimated figures. Record them as shown in column 7 of figure 9.

The floods for 1933–50 have now been adjusted to the base period 1914–50.

PRELIMINARY FREQUENCY CURVE

Plot discharge against recurrence intervals for each gaging station. Plot historical data in accordance with the discussion on page 17. Plot on forms similar to those illustrated in figures 3 and 4 (Survey form 9–179a and 9–179b).

Draw a frequency curve as a curve of visual best fit; do not compute mathematically. Extend the curve as far as the data warrants. A preliminary-frequency curve for Big Piney Run near Salisbury, Pa., is shown as figure 11; this curve bends up at the outer end, but if the data are plotted to a logarithmic-discharge scale a straight-line curve will be defined.

THE MEAN ANNUAL FLOOD

The mean used for each frequency distribution is the graphical mean determined by the intersection of the visually best fitting frequency line with the line corresponding to the 2.33-year recurrence interval. The graphical mean is more stable and dependable than an arithmetic mean. This method of determining the mean gives greater weight to the medium floods than to the extreme floods with large sampling errors, and for this reason is not influenced adversely

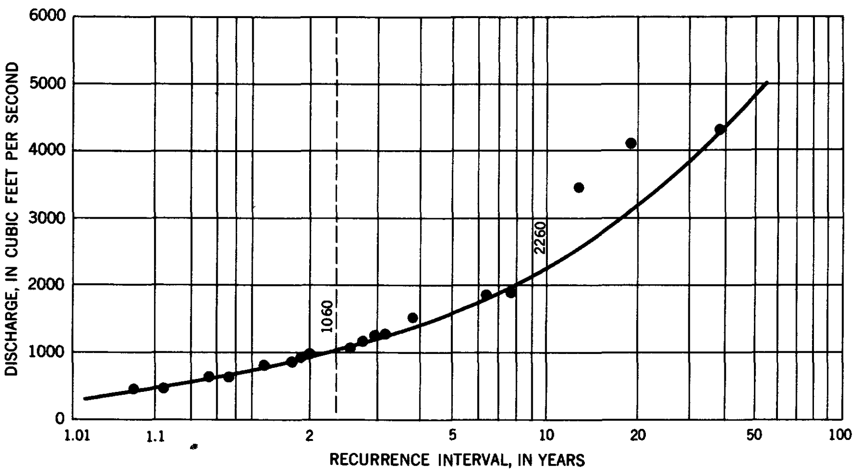

FIGURE 11.—Preliminary frequency curve, Big Piney Run near Salisbury, Pa.

by the chance inclusion or exclusion of a major flood, as is the arithmetic mean.

In determining the mean annual flood, the curve should be carefully fitted to the points plotting near the 2.33-year line. In the example, the mean annual flood for Big Piney Run is 1,060 cfs according to the curve of figure 11. (The arithmetic mean, computed from the 18 years of record, is 1,540 cfs.)

HOMOGENEITY TEST

The test setup requires a study of the 10-year floods as estimated at each station. Each 10-year flood should be divided by the mean flood to get the 10-year ratio, and an average of these ratios should be obtained. Tabulate for each station the length of record in years and the recurrence interval corresponding to a discharge equal to the average flood ratio times the mean flood. Data should be tabulated as shown in the following table:

Data for homogeneity test, base period 1914–50

No.	Stream	Drainage area (square miles)	Mean annual flood, $Q_{2.33}$ (cfs)	10-year flood, Q_{10} (cfs)	Ratio $\frac{Q_{10}}{Q_{2.33}}$	$Q_{2.33}$ ×1.70 (cfs)	T for Q of column 7 (years)	Period of record, adjusted (years)
1	2	3	4	5	6	7	8	9
1	Stony Creek	451	10,800	22,400	2.07	18,700	7	37
2	Conemaugh River	715	19,000	31,500	1.66	32,300	11	25
3	Conemaugh River	1,358	34,000	54,000	1.59	57,800	14	24
4	Kiskiminetas River	1,723	42,000	65,000	1.55	71,400	16	32
5	Little Conemaugh River	183	6,250	9,800	1.57	10,600	13	24
6	Blacklick Creek	390	11,500	18,800	1.64	19,600	12	37
7	Loyalhanna Creek	168	6,000	9,600	1.60	10,200	13	24
8	Loyalhanna Creek	265	7,400	11,800	1.60	12,600	13	28
9	Youghiogheny River	1,062	26,300	43,100	1.64	44,700	11	33
10	Youghiogheny River	1,362	31,500	58,100	1.84	53,500	8	33
11	Youghiogheny River	1,715	38,200	62,500	1.64	65,000	12	33
12	Casselman River	382	11,200	19,200	1.71	19,000	10	36
13	Big Piney Run	24.5	1,060	2,260	2.13	1,800	6	27
14	Laurel Hill Creek	121	4,850	7,400	1.53	8,250	17	37
15	Green Lick Run	3.07	230	500	1.79	476	8	30
Average ratio					1.70			

The adjusted period of record, column 9, is the number of years of actual record plus one-half the number of years the record was extended.

A set of test curves has been devised that shows within what range of recurrence intervals an estimate of a 10-year flood should be for a specified length of record; a range of 2 standard deviations is allowed. It is appropriate to base the test on the 10-year flood, because this is the longest recurrence interval for which many records will give dependable estimates. The test curves may be drawn on a standard chart, such as Survey form 9–179a, or a chart prepared from data given in the table under the section on "Homogeneity test" page 29. Plot the values of T_L and T_U.

As an example, data from columns 8 and 9 of the preceding table have been plotted on the chart as shown in figure 12. All points plot within the limits, therefore the records probably are acceptably homogeneous, and records from the 15 stations may be grouped together to define a regional flood-frequency curve. Had some points plotted outside the limits, it would have been necessary to separate the records into two or more groups and define a separate frequency curve for each.

COMPUTATION OF MEDIAN FLOOD RATIOS

The preliminary frequency curves (fig. 11) for all stations in a homogeneous region should be assembled. The ratios of several floods of different recurrence intervals to the mean annual flood should be tabulated for each station. Select only enough recurrence intervals to define the curve. Tabulate the flood ratios as shown in the following table. The ratios are easily computed as recorded, by slide-rule division.

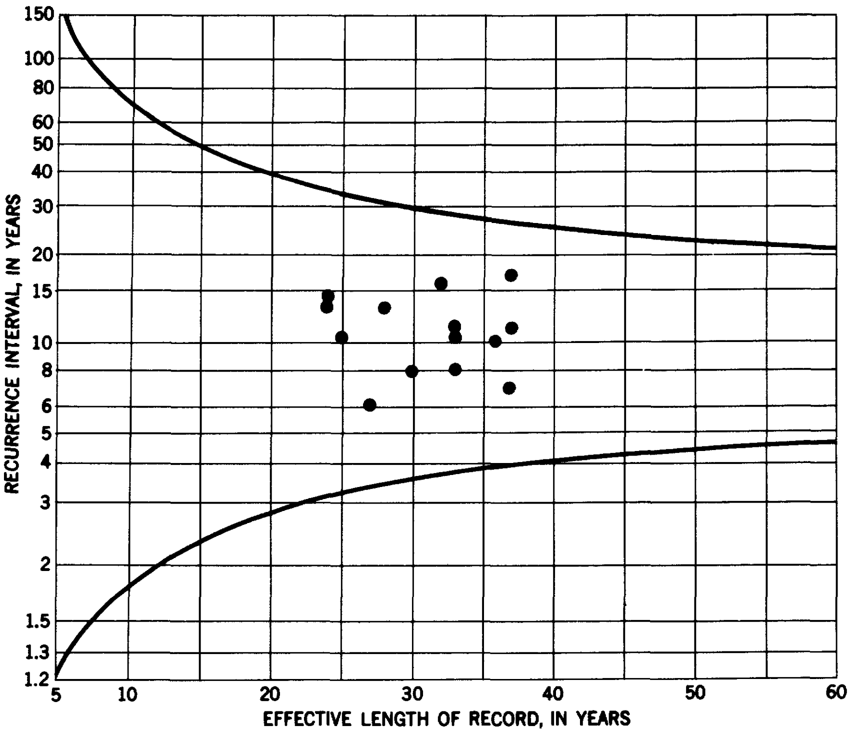

FIGURE 12.—Homogeneity test chart.

Median flood ratios, 1914–50 period

Station	Recurrence intervals, in years					
	1.1	1.5	5	10	20	50
1	0. 49	0. 75	1. 46	1. 93	2. 55	3. 03
2	. 57	. 78	1. 36	1. 74	2. 18	2. 73
3	. 54	. 79	1. 32	1. 55	1. 79	2. 09
4	. 57	. 80	1. 33	1. 62	1. 94	2. 48
5	. 56	. 79	1. 34	1. 63	1. 97	2. 52
6	. 49	. 78	1. 37	1. 65	1. 93	2. 28
7	. 50	. 75	1. 34	1. 70	2. 08	2. 57
8	. 45	. 76	1. 39	1. 76	2. 21	3. 02
9	. 69	. 83	1. 37	1. 80	2. 38	3. 00
10	. 62	. 81	1. 39	1. 77	2. 28	2. 92
11	. 58	. 78	1. 32	1. 67	2. 10	2. 90
12	. 63	. 80	1. 36	1. 80	2. 39	2. 91
13	. 47	. 74	1. 47	2. 13	3. 02	4. 57
14	. 46	. 76	1. 35	1. 64	1. 91	2. 25
15	. 54	. 75	1. 35	1. 73	2. 22	2. 80
Median	0. 54	0. 78	1. 36	1. 73	2. 18	2. 80

After tabulating the flood ratios, compute median ratios, as tabulated in the bottom line of the table. This median is the midvalue of an odd number of events or the mean of the two central values of an even number of events.

DEFINITION OF REGIONAL FREQUENCY CURVE

Each median flood ratio should be plotted to its corresponding recurrence interval on a frequency chart and an average frequency curve should be drawn. The regional frequency curve for the Youghiogheny and Kiskiminetas basins, based on the period 1914–50, is shown as figure 13.

The regional frequency curve, showing flood discharge in ratio to the mean annual flood, is based on all significant discharge records available and represents the most likely flood-frequency values for all areas in the region.

Adjustment from short to long period.—If more than one base period has been selected, the above process must be repeated for each period. If 2 base periods have been selected, 3 computations must be made:

1. Compute a regional frequency curve, as above, for the short period, using data for all stations.
2. Compute a regional curve for the long period, using records from all stations that are complete or have been filled in to the long period.
3. Compute a regional curve for the short period for only those stations that have records for the long period.

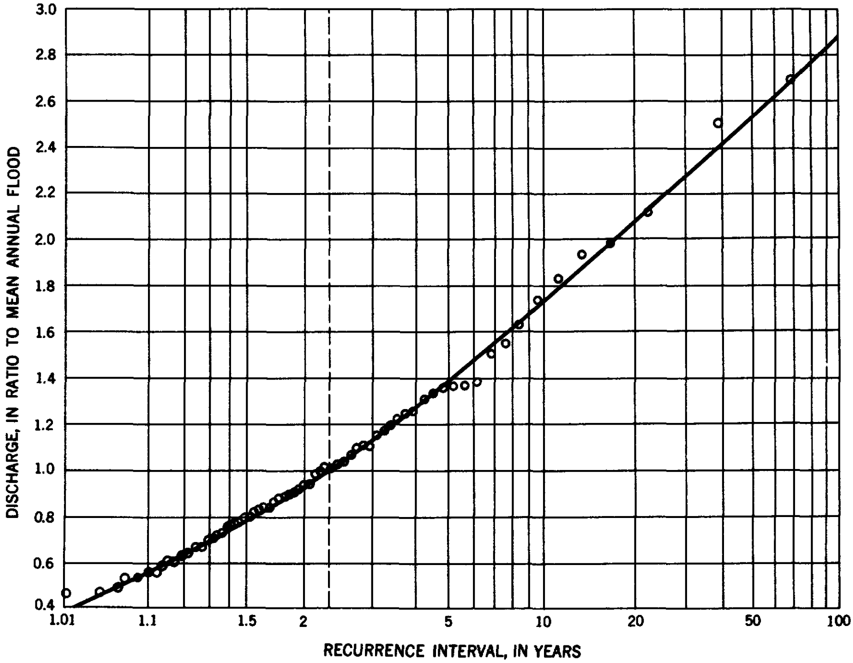

FIGURE 13.—Regional frequency curve.

Adjust the short-term curve computed in 1. to the long-term period by the ratio of the other two curves; that is, the frequency curve adjusted to the long-term period $=1.\times 2./3.$ This can easily be done by adjusting 6 or 8 points such as ratios for the 1.1-, 1.5-, 5-, 10-, 20-, and 50-year recurrence intervals.

If the short period curves extend to the 50-year interval, the long-period curves extend to the 100-year interval, and the adjustment below 50 years is appreciable, the 100-year point on the adjusted curve should extend the same amount above the 50-year point because the two are on the long-period curve.

ESTIMATION OF MEAN ANNUAL FLOOD

In order to apply the regional flood-frequency curve to an ungaged drainage basin estimate the mean flood for that basin. This involves a correlation analysis of the observed mean floods with drainage-basin characteristics.

If it is assumed that the region is hydrologically homogeneous, the factors which affect the mean flood are size, topography, shape of the drainage basin, and channel storage. Of these size is the most important, and the factor most readily available. Measuring the other factors is more difficult, and it may be impossible unless good topographic maps are available. Channel storage undoubtedly has an important effect, but cannot be directly measured.

For practical engineering use, a correlation of mean flood with drainage area may suffice. This simple correlation may require more than one curve for the region under study but eliminates tedious and lengthy computations that often may not be practicable. Mean annual-flood discharge as ordinate and drainage area as abscissa generally are plotted to log-log scales. A mean curve should be fitted graphically to the points. The mean annual-flood discharge for any stream in the region can be selected from this curve if the size of its drainage area is known.

A curve showing the variation of mean annual flood with drainage area is shown as figure 14. Data plotted on this figure were taken from columns 3 and 4 of the table shown in the section on "Homogeneity test".

Instead of plotting a point representing the discharge, plot a range line using the chance range in the mean annual flood based on the length of record. An average curve passing through all the range lines may be considered as the best determination of the relation curve. A discussion of this range line is given in the section " 'Working range' of mean annual floods," and figure 25.

Adjustment from short to long period.—If more than one base period has been selected, the mean annual floods derived from the short period must be adjusted to the long period before plotting. This

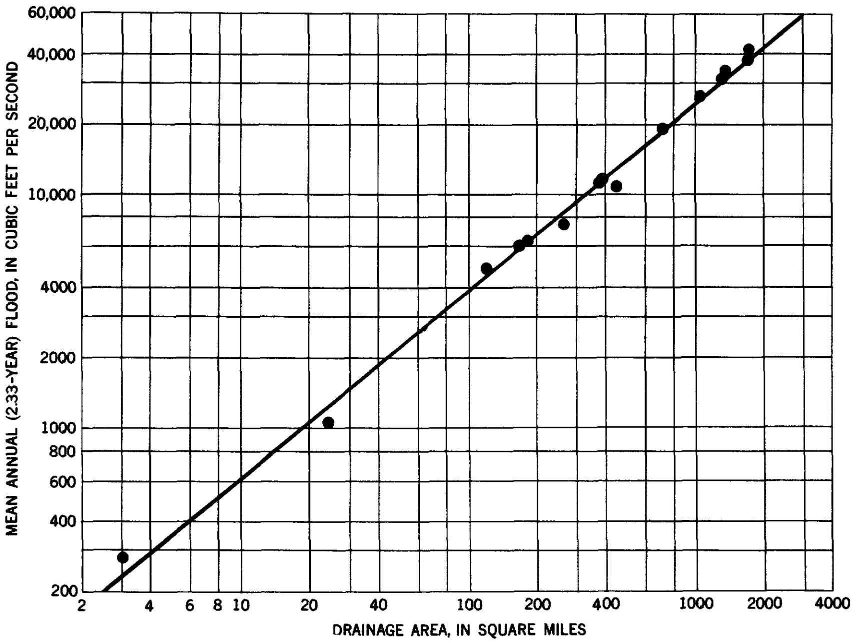

FIGURE 14.—Variation of mean annual flood with drainage area.

adjustment may be made by computing the average ratio determined from the 2.33-year floods computed for the two periods for the same gaging stations.

Correlation with physical factors.—Correlation of the mean annual flood with basin or meteorologic factors, other than the size of the drainage basin, can be made either by using statistical correlation techniques or by using graphical methods. Techniques of multiple correlation are covered in most textbooks on statistics. Results of correlation studies can be shown in entirely graphical form, such as a set of curves covering all the necessary relations. This has been the method used in most of the regional studies made so far. It is the simplest method of applying the results to ungaged areas as far as users of the report are concerned.

Results may be shown in the form of nomographs, as was done in a study for western Washington. (Bodhaine and Robinson, 1952, p. 28, 29). Or they may be shown in the form of equations, as was done for Connecticut (Bigwood and Thomas, 1955, p. 15). Equations are less convenient for the users; they present an opportunity for serious error in that they are apt to be used far beyond the limits of the data from which they were originally developed.

Regional frequency studies have been made in about half of the States; the factors that significantly correlate with the mean annual flood are: size of the drainage area (by far the most significant); slope of the principal streams; channel storage; area of lakes, ponds, and swamps; mean altitude; mean distance to outlet; annual runoff; basin lag time; soil types; and degree of urbanization.

Use of short-term records.—A long-term gaging station record is necessary for completely satisfactory flood frequencies. However, there are situations in which information is wanted in a brief time. In such situations, obtaining a record of all flood peaks above a base can be useful, provided there is a long-term record of a station nearby. The problem involves estimating of the mean annual flood; discharges of the rarer floods can be estimated from an appropriate regional frequency curve, such as the one shown as figure 13.

From a reasonably short record (2-6 years), barring an uncommon sequence of drought years, a short partial-duration series of floods can be developed, at least in the lower range. Graphical comparison of such discharges at a short-term station with the corresponding discharges for the same period at a station with a long record, where the mean annual flood is known, would permit estimation of the mean annual flood at the short-term station.

Partial-duration series, 1946–47

Order	Discharge in cubic feet per second		Order	Discharge in cubic feet per second	
	Short-term station (Olentangy River near Delaware, Ohio, 1946–47)	Long-term station (Scoito River near Columbus, Ohio, 1913, 1921–48)		Short-term station (Olentangy River near Delaware, Ohio, 1946–47)	Long-term station (Scoito River near Columbus, Ohio, 1913, 1921'48)
1	9, 160	31, 600	6	5, 190	13, 600
2	8, 640	20, 500	7	5, 190	11, 700
3	6, 060	19, 600	8	4, 200	11, 700
4	5, 840	19, 600	9	3, 840	11, 000
5	5, 730	14, 000			

The data for corresponding order numbers are plotted against each other in a logarithmic chart, as shown in figure 15. A straight line was drawn to average the plotted points. The mean annual flood for the long-term record is 27,500 cfs. The estimate for the short-term station is 10,000 cfs. The record on Olentangy River near Delaware Ohio, is in fact 36 years long, and the mean annual flood is 9,000 cfs, a 10 percent difference from the foregoing estimate.

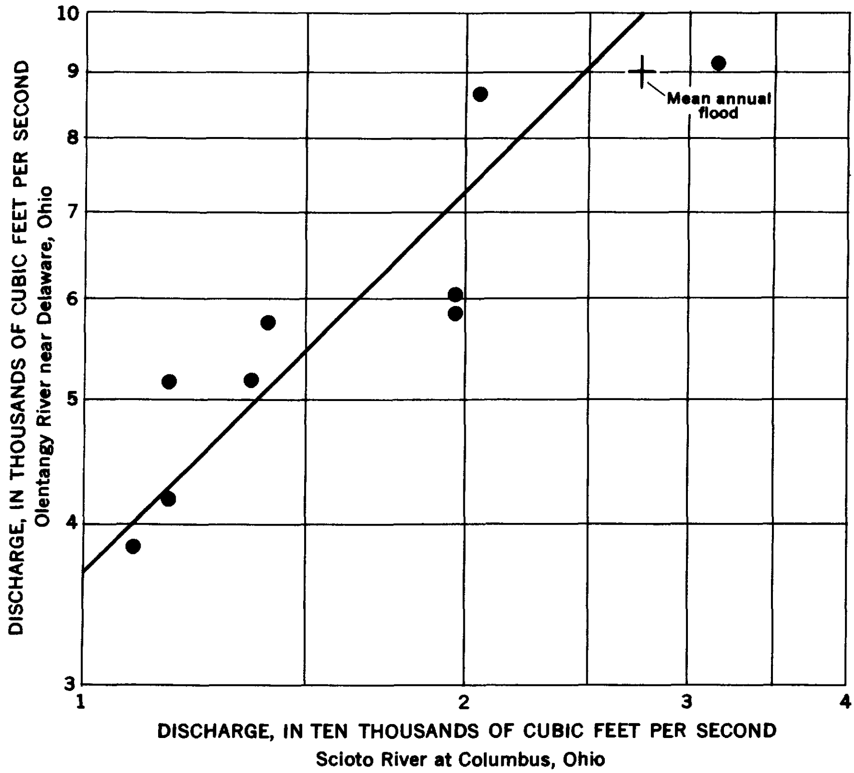

FIGURE 15.—Determination of mean annual flood from a short record.

DEFINITION OF A FREQUENCY CURVE

To define a flood-frequency curve at any place within the study region, the mean annual flood is found from the curve relating this factor to the drainage area (fig. 14). The mean annual flood is multiplied by the flood ratios previously computed (fig. 13) to obtain the discharge corresponding to a selected frequency. A complete frequency curve can be produced by plotting discharges for several frequencies and drawing the curve they define.

An example of a flood-frequency curve defined for any 100-square-mile drainage basin in the Youghiogheny or Kiskiminetas River basins in Pennsylvania is presented as figure 16.

SUMMARY OF PROCEDURE

The step-by-step procedure for making a regional flood-frequency analysis follows:

1. Tabulate flood data for all gaging stations in the region having a record of 5 years or more. List the maximum annual floods only. Leave space for tabulating adjusted data.
2. Prepare a bar graph and select the base period for study; generally this will be the period of the longest record.
3. Adjust all the records to the base period.
4. Number the floods for each station in the order of magnitude, the greatest flood is number 1.
5. Compute the recurrence intervals.
6. Plot the discharge against the recurrence intervals and draw frequency curves, one for each station.
7. Test for homogeneity.
8. Compute the median flood ratios.

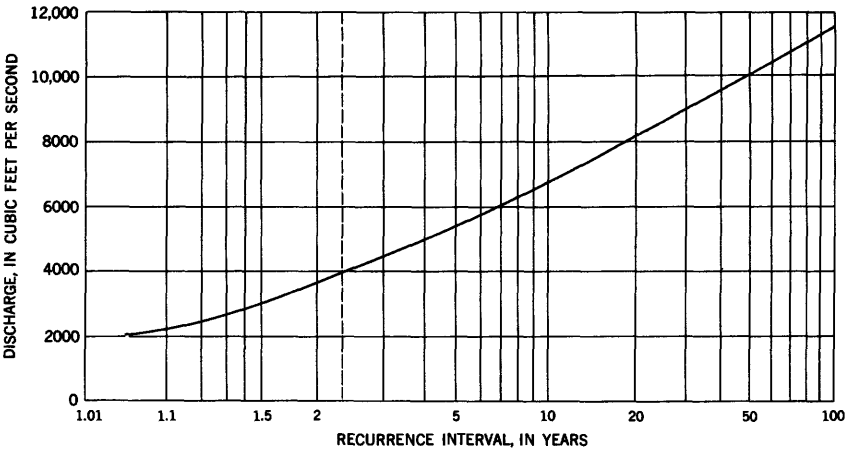

FIGURE 16.—Flood-frequency curve for a 100-square-mile drainage basin.

9. Plot the median flood ratios and draw a regional frequency curve.
10. Plot the mean annual floods against the drainage areas. Draw a curve, or curves, to show tne relation applicable for the region.
11. Determine flood frequency for any place in the region from the curves of items 9 and 10.

SPECIAL TREATMENT FOR LARGE STREAMS

Large streams that traverse more than one flood-frequency region or hydrologic area may not fit into the general pattern, and special treatment is indicated. The frequency relation may follow one of the regional curves plotted for tributary streams, where the regional curve would be used. If it follows an independent pattern, a frequency curve may be plotted for the one stream. This is illustrated by the curve defined for the Mississippi River by Searcy (1955, fig. 12).

The mean annual floods for large streams may require separate curves derived from records for stations along those streams. The curves smooth the individual station records and are a means of interpolation between the stations, on the basis of either size of drainage area or river mileage. An example of such a curve for the Mississippi River, taken from Searcy (1955, fig. 11) is presented as figure 17.

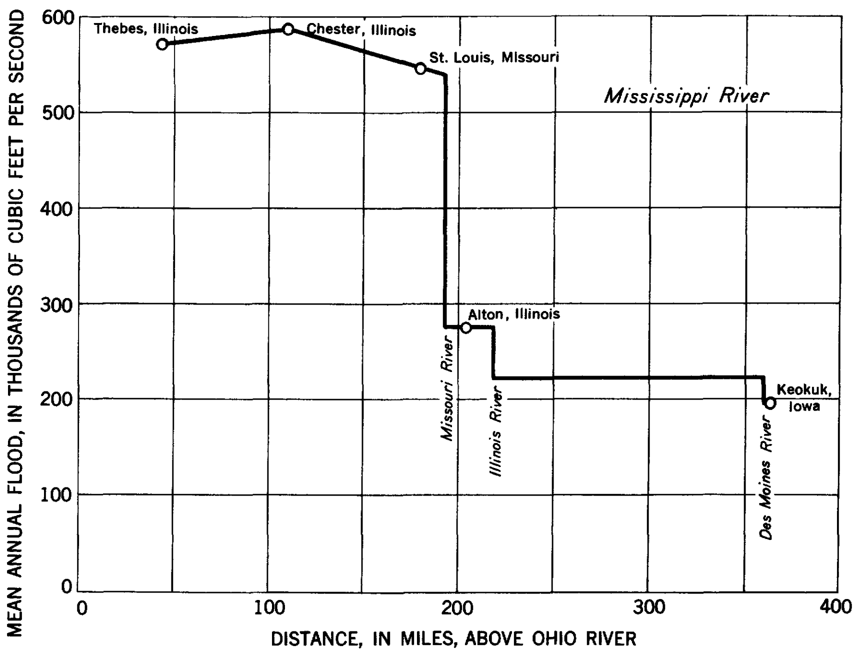

FIGURE 17.—Variation of mean annual flood with distance along stream.

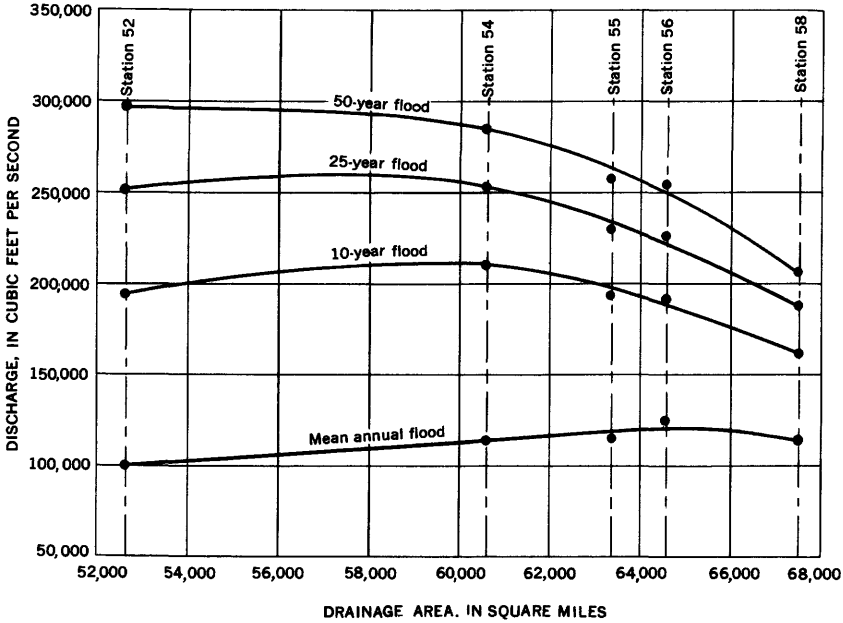

FIGURE 18.—Relation of flood frequencies to drainage area.

Some large streams show a change in both the flood-frequency relation and the mean annual flood at gaging stations along the stream. Separate curves for selected floods may be defined and used for interpolation along the stream. An example of this treatment has been made for the Red River in Louisiana by Cragwall (1952, fig. 22) and is presented as figure 18.

Regional flood-frequency reports have been prepared by methods described in this report and have been published by the Geological Survey or cooperating agencies. Knowledge of techniques may be gained from a study of these reports. All are listed in "Selected references"; they are listed below in brief, for ready reference:

Alabama	(Peirce, 1954)
Columbia River basin.	(Rantz and Riggs, 1949).
Connecticut	(Bigwood and Thomas, 1955)
Delaware River basin.	(Tice, 1958)
Florida	(Pride, 1957)
Georgia	(Carter, 1951)
Illinois	(Mitchell, 1954)
Iowa	(Schwob, 1953)
Kentucky	(McCabe, 1958)
Louisiana	(Cragwall, 1952)
Minnesota	(Prior, 1949)
Missouri	(Searcy, 1955)
Montana, eastern part.	(Berwick, 1958)
Nebraska	(Furness, 1955)
North Carolina	(Riggs, 1955)
North and South Dakota.	(McCabe, 1957)
Ohio	(Cross, 1946)
Washington, western part.	(Bodhaine and Robinson, 1952)
Youghiogheny and Kiskiminetas River basins, Pennsylvania and Maryland.	(Noecker, 1952)

PLOTTING POSITIONS IN FREQUENCY ANALYSIS

By W. B. Langbein

The formula $\frac{n+1}{m}$ for recurrence intervals has become nearly standard in Survey reports on flood and drought frequency. The first use of this formula is ascribed to Kimball (1946), but except for Survey reports and a few papers in statistical literature, this formula has been little used. It, therefore, may be unfamiliar to many of the readers of Survey reports, and some explanation is desirable concerning the nature of the formula in relation to others mentioned in flood literature.

For simplicity this explanation will treat only the highest flood in an annual series for which the several formulas for plotting positions differ most.

Consider first, for example, the question: What is the chance that a 50-year flood will be the maximum in a 10-year period? A 50-year flood has a probability of 0.02 of being equaled or exceeded in any year and a probability of 0.98 or not being equaled or exceeded in any year. A flood of this magnitude can be the maximum in a 10-year annual series only if all 10 values are equal to or less than the 50-year flood. Hence 0.98^{10} represents the probability that a 50-year flood will be the greatest in a 10-year period. This logic is general and is independent of the nature of the flood-frequency graph. The general formula therefore is:

$$z=(1-p)^n \tag{1}$$

$$p=\frac{1}{T} \tag{2}$$

in which z represents the probability (percent chance) that a flood of recurrence interval T or probability p will be the maximum in an n-year period.

Assume that several records of floods are available, each of n years. The top floods in each list would vary greatly in their magnitude; some would be of 5-year size, some of 50-year size, and so on. The ratio z given by equations 1 and 2 may also be viewed as representing the relative proportion of floods less than the T-year flood among the top floods in the groups of n floods. If there were several lists of n floods, one could verify equation 1, but there is only one list of floods of n years and only one maximum of this list. The problem is to assign a representative plotting position to this single experience.

A single value offers no opportunity for comparison or for any other measure of central tendency. The measures of central tendency generally are the median, the average, or the mode, as defined in any elementary statistics text. The only possible course is to accept one of these measures of central tendency; the median, the arithmetic mean, or the mode. The selection of one of these is the basis for the several formulas for "plotting positions."

MEDIAN

This assumption was introduced by Beard (1943). The value of z is set equal to 0.50. Plotting positions are therefore as follows: $p=1-0.50^{\frac{1}{n}}$ or the reciprocal where recurrence intervals are desired. For recurrence intervals the solution converges toward $0.5+1.443^{n}$ as an approximate answer.

MEAN

In discussing plotting positions Gumbel (1945b, p. 70) says, "We may conclude from the distribution of the largest value that the mean surpasses the median, and the occurrence-interval surpasses 1.433^{n}. Therefore, the mean of the largest value cannot be used for the determination of the occurrence-interval." In his report, Gumbel does not consider the mean of the probabilities of the largest values. However, in a later article Gumbel and von Schelling (1950) discuss the distribution of exceedences from which Chow (1953) show the mean number of exceedences in N future trials to be $N=\frac{m}{n+1}$, or simply $\frac{m}{n+1}$ for the probability of the mean number of occurrences. This is the formula for plotting position used by the Survey. This formula can be derived for the special case of the highest value, $m=1$, simply by direct integration of equation 1 between the limits of $z=0$, and $z=1$. The integration is as follows:

$$\bar{p}=\frac{\int_0^1\left(1-z^{\frac{1}{n}}\right)dz}{\int_0^1 dz}=\frac{1}{n+1}$$

for the plotting position of the highest value in terms of its probability and $n+1$ in terms of recurrence interval.

This formula for plotting position can also be stated as follows: A list of n events contains $(n+1)$ class intervals. A future event has equal probability of falling within any of these intervals. Because the highest value represents the bottom of the top class interval, it represents an event which has a probability of $\frac{1}{n+1}$ of being equaled or exceeded.

MODE

This is the basis of the plotting positions employed by Gumbel in his original treatment of the problem. Equation 1 has no mode, hence a modal position can only be considered in relation to a distribution of the flood magnitudes. Combining equation 1 with the equation for the distribution of extreme values $1-p=e^{-e^{-y}}$, in which y is a linear function of the discharges, leads to the equation $z=(e^{-e^{-y}})^n$. Setting $\frac{d^2z}{dy^2}$ equal to zero for the mode gives: $p=1-e^{\frac{1}{n}}$ for the probability of the mode. For values of n greater than 10, the value of $1-e^{\frac{1}{n}}$ equals $\frac{1}{n}$ very closely. Hence, where there is a long record, the modal plotting position in terms of recurrence intervals equals n the number of years in the record. This is the procedure in the California method and is implied in duration-curve work.

How does the Hazen method (see p. 15) fit into the scheme of plotting position? Although it is not defined by reference to equation 1, it is possible to determine where it fits. According to the Hazen formula, the highest flood has a plotting position that corresponds to $2n$ in terms of recurrence intervals and $\frac{1}{2n}$ in terms of probability p. Therefore $z=\left(1-\frac{1}{2n}\right)^n$ whereby $z=0.60$ for most values of n.

Comparison of the results of the preceding plotting positions for the highest flood

Method	Designation	Plotting position	Percentage of future floods that will be less than plotting position [100−(1−z)]
Hazen		$\frac{1}{2n}$	60
Beard	Median	$1-0.50^{\frac{1}{n}}$	[1] 50
Geological Survey	Mean	$\frac{1}{n+1}$	40
Gumbel	Modal	$1-\frac{1}{e}^{\frac{1}{n}}$	37

[1] By definition.

The final column lists the percentage of the floods in the future that will be less than the given plotting position. The Hazen method, which gives the highest position, will be fallen short of by 60 percent of future floods, whereas the Gumbel modal position will be less than 37 percent of the future floods.

Also note that the "mean" position lies in between those of the median and the mode, which may seem anomalous, because in most examples it is the median that lies between the mean and the mode. However, the mode in this table applies to a somewhat different situation than the median and mean.

The selection of a plotting position for the top flood of a list is like choosing a stand on a political question; there is a wide range of reasonable positions. But uniformity is essential if reports on flood frequencies are to be comparable. The term $(n+1)$ for the recurrence interval of the highest flood of record was adopted for use in flood-frequency reports of the Geological Survey because it corresponds to a mean probability and because it is simple to apply.

CHARACTERISTICS OF FREQUENCY CURVES BASED ON A THEORETICAL 1,000-YEAR RECORD

By M. A. Benson

This report contains a summary of the procedures followed and the results obtained in analyzing a "perfect 1,000-year record" of maximum annual peak floods.

The object of the investigation was to study possible variations in frequency curves computed from short periods of record, each taken from a long-term record whose frequency characteristics were exactly determinable. There is no "theoretically correct" frequency distribution of flood events, although claims have been made for one or another system either based on empirical evidence or (with necessary assumptions) on statistical theory. For this reason, we are not sure of the "true" shapes of the frequency curves of even the longest flood records.

It was therefore decided to start from an assumed list of 1,000 peak floods, defining exactly a simple-frequency curve. Separate parts of the array could be analyzed independently and the results compared with the known characteristics of the base curve. This could demonstrate the variations, due to chance alone, in frequency curves from short records.

BASE DATA

The data forming the basis for this study are a list of 1,000 figures of varying magnitudes between the arbitrary limits of 9,910 and 380. A partial listing of the base data is shown in the following table:

Flood frequency data, 1,000-year record (partial listing)

Order No. (1−1,000) (m)	T (years) $=\frac{1,000+1}{m}$	Discharge (cfs)	Year (1−1,000) (m)	Discharge (cfs)	10-year period		25-year period		50-year period		100-year period	
					Order No.	T (years)	Order No.	T (years)	Order No.	T (years)	Order No.	T (years)
1	2	3	4	5	6	7	8	9	10	11	12	13
1	1,001	9,910	957	1,220	9	1.22	23	1.13	48	1.06	95	1.06
2	500	9,180	456	2,980	7	1.57	15	1.73	33	1.55	62	1.63
3	334	8,730	167	4,280	2	5.50	4	6.50	10	5.10	23	4.39
4	250	8,410	348	3,370	6	1.83	13	2.00	27	1.89	51	1.98
5	200	8,190	75	5,180	1	11.0	1	26.0	3	17.0	11	9.18
6	167	7,990	237	3,840	4	2.75	6	4.33	16	3.19	32	3.16
7	143	7,820	977	1,040	10	1.10	24	1.08	49	1.04	98	1.03
8	125	7,670	276	3,660	5	2.20	9	2.89	20	2.55	39	2.59
9	111	7,540	175	4,220	3	3.67	5	5.20	11	4.64	25	4.04
10	100	7,420	868	1,690	8	1.38	21	1.24	46	1.11	89	1.13
11	91.0	7,310	146	4,430	2	5.50	3	8.67	8	6.38	20	5.05
12	83.4	7,210	597	2,550	8	1.38	17	1.53	37	1.38	73	1.38
13	77.0	7,120	410	3,130	6	1.83	14	1.86	30	1.70	57	1.77
14	71.5	7,030	137	4,500	1	11.0	2	13.0	7	7.29	19	5.32
15	66.7	6,960	305	3,540	5	2.20	12	2.17	25	2.04	47	2.15

The data used were obtained as follows:

1. An arbitrary straight line was drawn on the frequency graph of figure 19, with an arbitrary discharge (ordinate) scale.
2. One thousand values of discharge were read from this graph for abscissa values of $T=\frac{n+1}{m}$, where n equals 1,000, and m, the order number, ranges from 1 to 1,000. The computed values of T and the corresponding discharges are shown in columns 2 and 3 of the table.
3. Numbers ranging from 1 to 1,000, representing m, were written on slips of papers, and these were thoroughly shuffled, then withdrawn one at a time in random order, as shown in column 4 of the table. The corresponding discharges are shown in column 5.

The random variation from an "infinite" population could have been approximated by returning each slip to the pack and reshuffling before drawing another. However, the infinite range in discharge existing in an infinite population could not be duplicated, so that the resultant arrray would have uncertain properties. The objective was to obtain a random distribution of floods in a postulated 1,000-year record.

The list of 1,000 peak floods is similar to what might happen in nature, assuming that in a 1,000-year period a "perfect" distribution can occur. In this distribution one 1,000-year flood; two 500-year

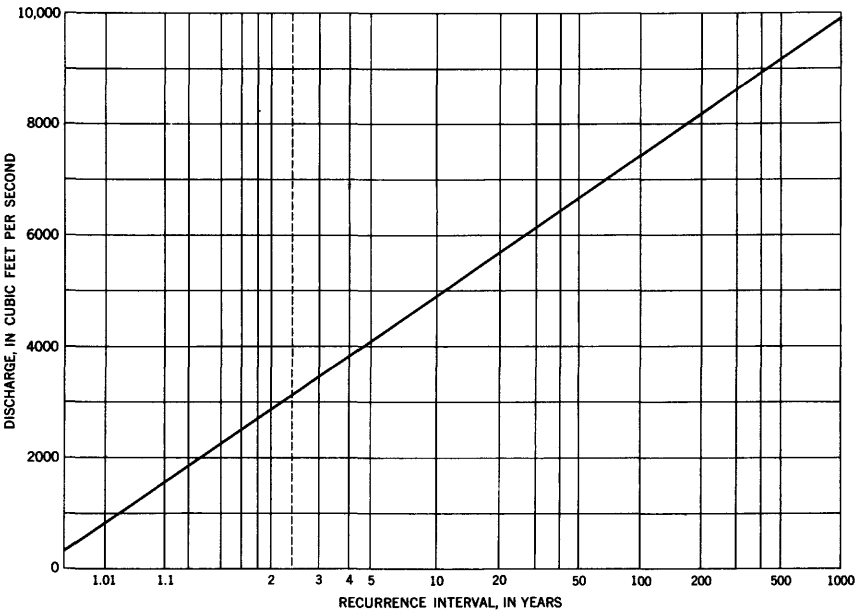

FIGURE 19.—Theoretical frequency curve.

floods; three 333-year floods; and so on occur throughout the entire range. They are in random order, and short consecutive periods of this array are similar to what might be found in short periods of record in nature. Analysis of the frequency curves computed from short periods may be compared with the known characteristics of the basic long-term frequency curve. Since the order of the items was established by random drawing, any differences from the basic curve are due to chance alone.

MEAN ANNUAL FLOOD

According to the theory of largest values (Gumbel, 1945a), the mean of all the annual floods has a value corresponding to the flood of 2.33-year recurrence interval. The arithmetic mean of the 1,000 items shown in column 5 of the preceding table is 3,064, which is the same as the value at 2.33-years on the base curve of figure 19.

SHORT PERIODS OF RECORD

PERIODS USED

The base data were divided into "records" of various lengths for purposes of analysis; for example, in constituting twenty 50-year "records" the first 50 on the random list was the first "record" and so on. Division was as follows:

1. One hundred consecutive 10-year periods:
   Items 1–10, 11–20, 21–30—991–1,000.
2. Forty consecutive 25-year periods:
   Items 1–25, 26–50—976–1,000.
3. Twenty consecutive 50-year periods:
   Items 1–50, 51–100—951–1,000.
4. Ten consecutive 100-year periods:
   Items 1–100, 101–200—901–1,000.

Each of the 170 periods was analyzed separately, as though it were an independent record of 10, 25, 50, or 100 years. Recurrence intervals were computed and plotted and individual frequency curves were drawn for each.

DRAWING OF FREQUENCY CURVES

In drawing the frequency curves, the same procedures were followed as when actual station records are being studied:

1. The curves were drawn by eye as the graphical curves of best fit.
2. They were drawn as straight lines or curves depending on the trend of the points.
3. The curves were drawn to average rather than to follow individual points, the object being to avoid sharp breaks or bends.
4. The curves were kept as close as was reasonably possible to the plotted points in the vicinity of the 2.33-year line.

5. **Floods of high recurrence interval are likely to occur within a short period of record. For this reason, the highest points were not given much weight in defining the curves, if they were to the left and above the trend line through the remaining points. If the higher points lined up consistently with each other and with the lower points, they were followed although the resulting curve bent upward to the left.**

It is not possible to avoid the personal factor in drawing curves fitted visually as described here. However, this is believed the most practical method and gives the best results. A general confirmation of the method was furnished by the following means. Values of the 2.33 (mean annual), 10-, 25-, 50-, and 100-year floods were taken from the individual curves (except that the 50-year flood was the highest taken from the 10-year curves). The averages of these values compared with the known true values (from the base curve, fig. 19) are shown in the following table:

Comparison of true values with averages of computed values of mean annual flood

Length of record (years)	Mean annual flood (cfs)	10-year flood (cfs)	25-year flood (cfs)	50-year flood (cfs)	100-year flood (cfs)
10	3, 100	5, 060	6, 160	6, 970	------
25	3, 110	4, 940	5, 960	6, 720	7, 440
50	3, 100	4, 900	5, 900	6, 660	7, 390
100	3, 080	4, 860	5, 940	6, 770	7, 570
True value	3, 064	4, 860	5, 860	6, 680	7, 420

Most of these averages are within small percentages of the known values. The average values of the 10-, 25-, and 50-year floods as derived from 10-year records are between 4 and 5 percent higher than the true values. This may seem excessive and might demonstrate that curves are being drawn too high. However, a review of the individual frequency curves shows that this is not a general tendency.

There is always some uncertainty as to whether to draw the curve to follow the upper points, if it is necessary to bend the curve upward to do so. This study can give no guidance for any single example which may be met in practice, but some frequency curves have such a tendency. The only way to determine whether such a graph represents a physical actuality or is merely chance, is to compare it with curves for other places in the region.

Some investigators prefer to fit frequency curves as straight lines by computations as outlined by Powell (1943). A serious objection to this procedure is that if a short period of record contains one or more long-term floods, the line is unduly influenced by these and may be drawn to the left of most of the other floods. To avoid this, some investigators advocate drawing a straight line through the lower points, giving little weight to the upper points. Where all points indicate definite curvature throughout the range, this procedure becomes wholly arbitrary and subjective. The greatest objection to the drawing of straight-line graphs, whether computed or by eye, is that it is first necessary to support the belief that they should be straight lines. There is no firm basis for rigid adherent to straight lines.

The individual frequency curves for 10-, 25-, 50-, and 100-year periods are shown as figures 20 to 23, respectively. All those based on a common length of period have been plotted together to show graphically the extent of variation obtained.

ARITHMETIC VERSUS GRAPHICAL MEAN ANNUAL FLOOD

The mean of the annual floods, or "mean annual flood," is an extremely important factor, used in correlation studies and in regional flood-frequency compilations. It is desirable to determine it from fairly short-term records.

In short-term records particularly, the arithmetic mean is affected considerably by the chance inclusion of one or more major floods. The Geological Survey uses the graphical mean annual flood to avoid this adverse condition. The graphical mean is the value determined by the intersection of the visually best fitting frequency curve with the mean line (the line corresponding to the 2.33-year recurrence interval). The graphical mean is more stable and dependable than the arithmetic mean. This method of determining the mean gives greater weight to the medium floods than to the extreme floods with large sampling errors. The resulting figure is no longer the mean of the annual floods, but is the "mean annual flood" by definition.

This investigation makes possible a practical comparison between values of arithmetic and graphical means, and a demonstration of the variation in the value of the mean with the length of the record.

The average of the arithmetic means of all the short-term periods is 3,064. The averages of the graphical means, as obtained from the individual short periods, vary as follows:

Length of record (years)	Number of records	Average of graphical means
10	100	3, 100
25	40	3, 110
50	20	3, 100
100	10	3, 080

These are within small percentages (less than 1.5 percent) of the true values, so that no gross errors are involved in using the graphical mean.

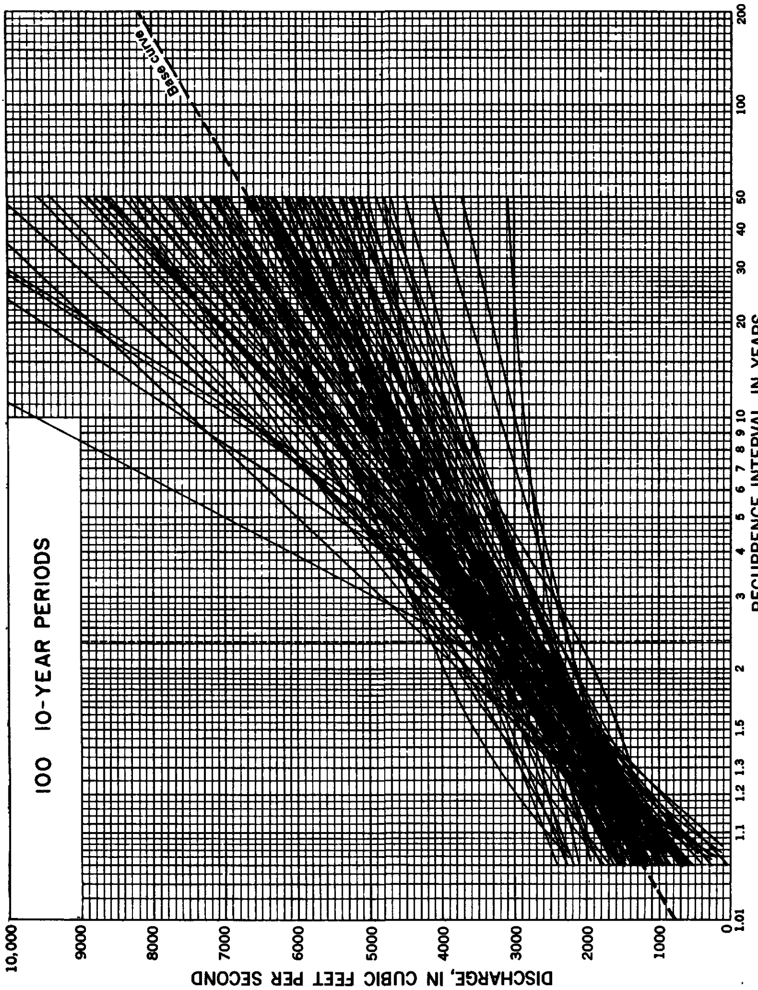

FIGURE 20.—Frequency curves for 10-year periods.

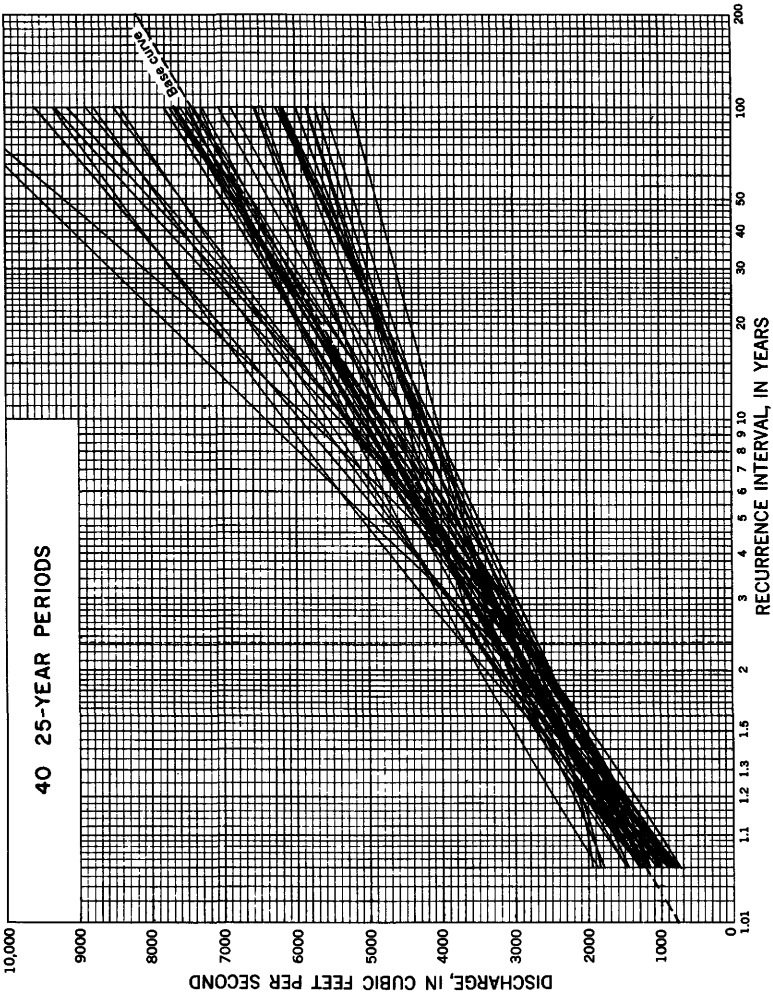

FIGURE 21.—Frequency curves for 25-year periods.

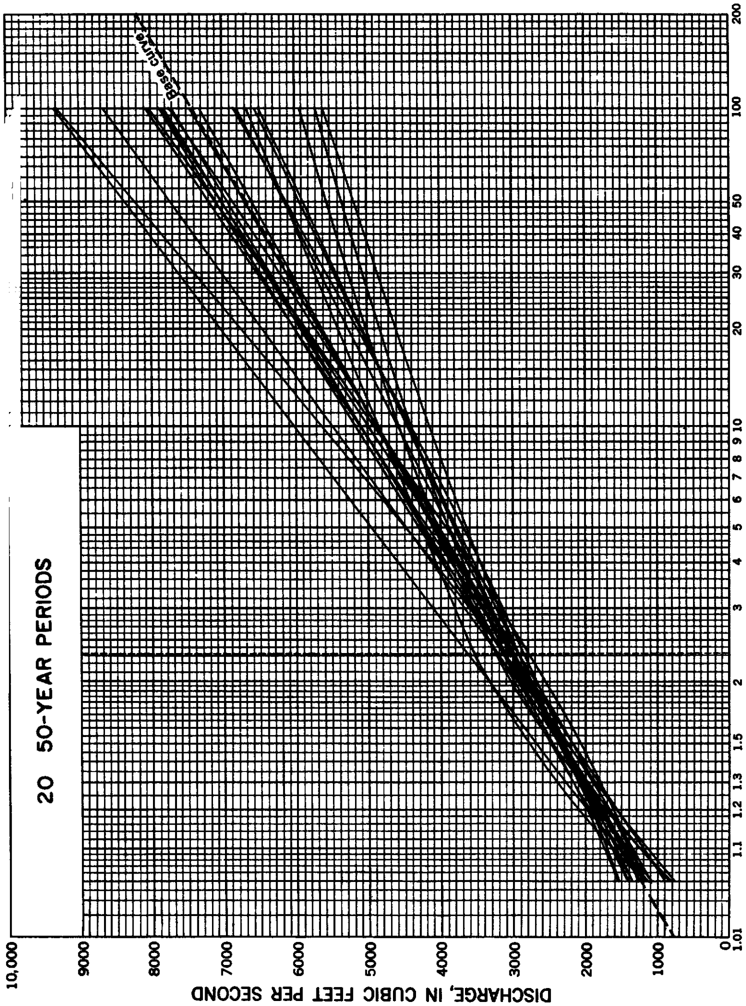

FIGURE 22.—Frequency curves for 50-year periods.

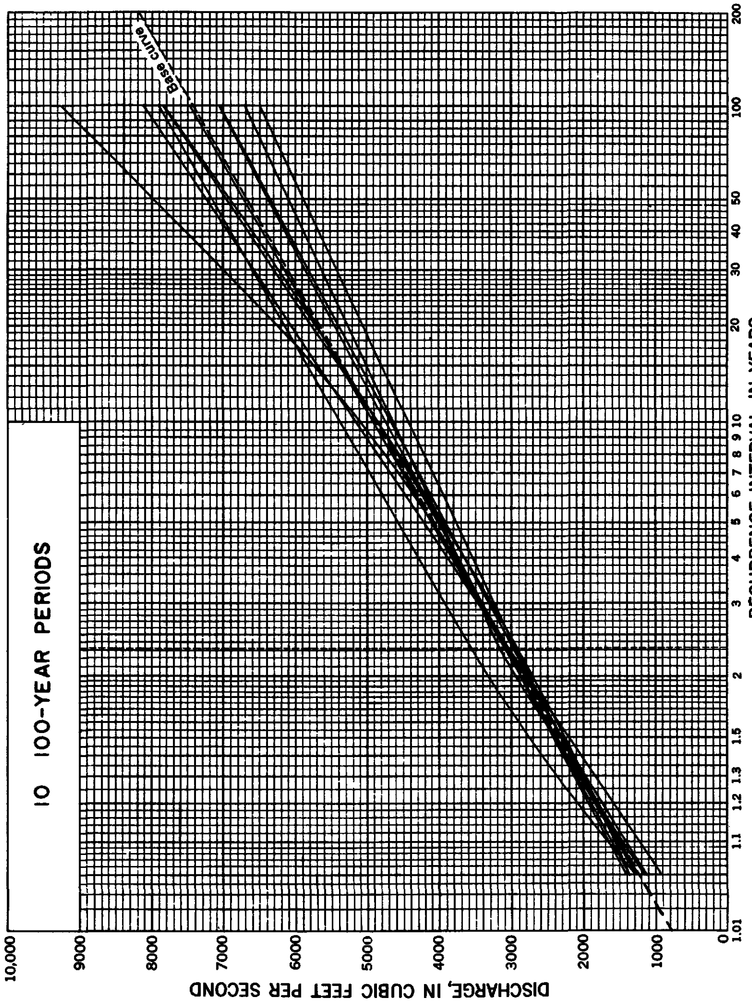

FIGURE 23.—Frequency curves for 100-year periods.

RELIABILITY OF MEAN ANNUAL-FLOOD VALUES

In statistical studies, the most favorable expectancy during 95 percent of the time is commonly used as the criterion for dependable results.

The extent of variation from the true mean of the mean annual floods (determined graphically) in this study, during 95 percent of the time, was 28 percent for 10-year periods, 14 percent for 25-year periods, 12 percent for 50-year periods, and about 5 percent for 100-year periods.

These percentages cannot be assumed to apply generally, because they will vary with the slope of the base-frequency curve. They are a general indication of what may be expected from short- and long-term records. The 10-year records will give a much wider range in determining the mean.

It is desirable to have some means of allowing for the range in value due to chance, and a method is hereby outlined which gives consideration to the length of the record from which the mean annual flood has been determined.

"WORKING RANGE" OF MEAN ANNUAL FLOODS

Based on the theory of extreme values, it is possible to compute, for conditions existing 95 percent of the time, and for any length of record, the range in recurrence intervals which would be found for the 2.33-year flood. The theoretical limiting curves for these values are shown as figure 24.

These curves are generally applicable. The extent to which results of this 1,000-year frequency study conform with the theoretical curves is shown by the plotted points on figure 24. These points represent the apparent recurrence intervals corresponding to the known mean annual flood of 3,064, as determined from the 170 individual frequency curves developed in the study. Theoretically, 5 percent of these points, or 8.5 points, should be outside the limits; actually, 7 of the points lie outside.

In regional flood-frequency studies, the relation between the mean annual flood and the drainage area (or other basin characteristics) is required and can be determined graphically by plotting one against the other. Instead of plotting a point, a range line may be plotted using the chance range in the mean annual flood based on the length of record. An average curve passing through all the range lines may then be considered as the best determination of the relation curve.

The method of determining the chance or working range of the mean annual flood is illustrated in figure 25. Curve *A* is an assumed frequency curve, based on a presumed record of 10 years. The curves

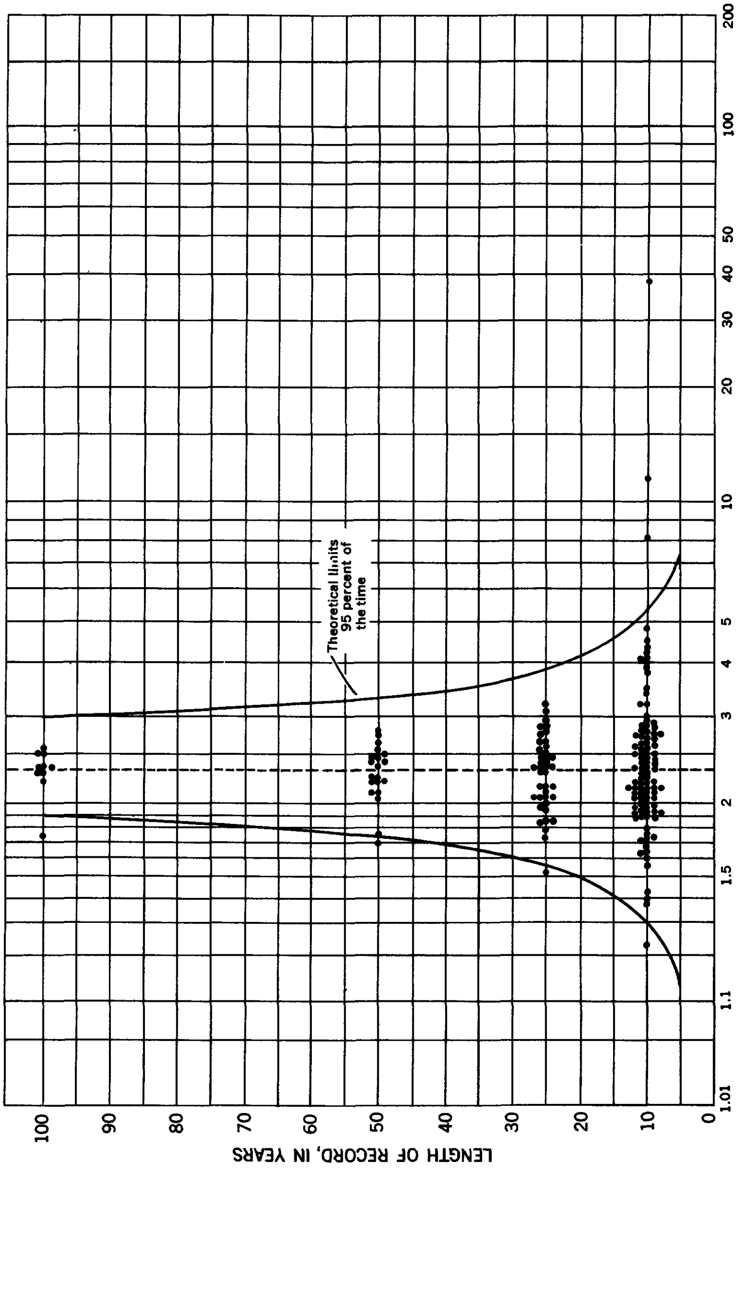

FIGURE 24.—Recurrence intervals of 2.33-year flood.

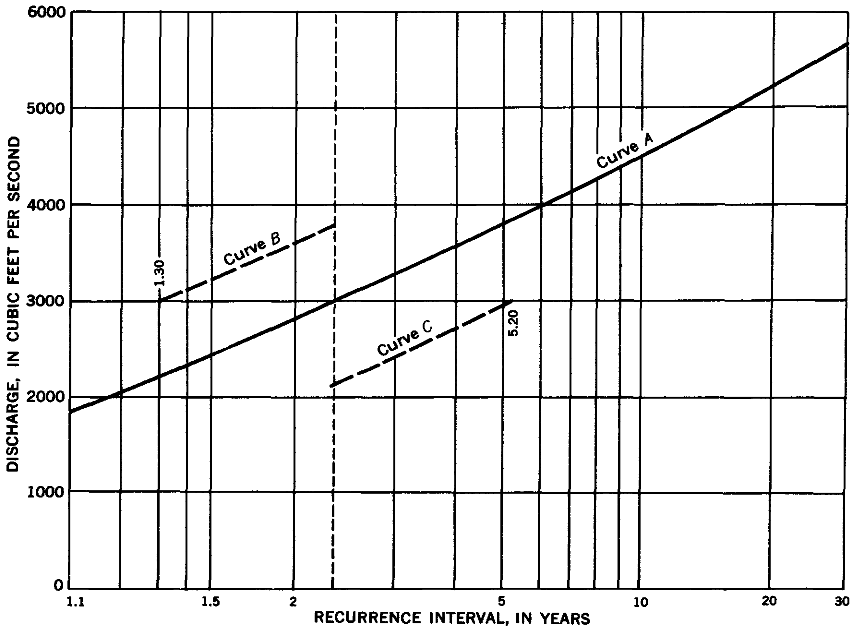

FIGURE 25.—Computing range for mean annual flood.

of figure 25 show that the recurrence interval for a mean annual flood based on 10 years of record might have an apparent recurrence interval of between 1.30 and 5.20 years (95 percent of the time). The mean annual flood, determined from curve *A*, is 3,000 cfs. Curves *B* and *C* are drawn parallel to curve *A* passing through 3,000 cfs at recurrence intervals of 1.30 and 5.20 years. These represent the extreme positions which curve *A* might take due to the chance variation in 10 years of record. (It is assumed that the slope would be parallel within these short segments.) These curves intersect the 2.33-year line at 2,100 and 3,800 cfs; this range is interpreted as showing that the actual value of the mean annual flood might, due to chance variation, lie anywhere between 2,100 and 3,800 cfs. The mean annual flood for this station record is plotted as a range line between these limits, against any other desired factor.

RELIABILITY OF FLOOD MAGNITUDES

A study was made of the maximum and minimum values of the 10-, 25-, 50-, and 100-year floods as determined graphically from records of various lengths, for 80, 95, and 100 percent of the time.

These results show, for example, that (19 out of 20 times) for these data, any 39-year record can define the 50-year flood within 25 percent of the true, long-term value.

Lengths of record necessary to come within 25 percent of the correct value 95 or 80 percent of the time

Magnitude of flood (T in years)	Length of record in years	
	95 percent of the time	80 percent of the time
2.33	12	--------
10	18	8
25	31	12
50	39	15
100	48	--------

Lengths of record necessary to come within 10 percent of the correct value 95 or 80 percent of the time

Magnitude of flood (T in years)	Length of record in years	
	95 percent of the time	80 percent of the time
2.33	40	25
10	90	38
25	105	75
50	110	90
100	115	100

In recent years, many crest-stage gaging stations have been established for the purpose of defining the flood potential of a region. The mean annual flood can be determined within 25 percent (95 percent of the time) by a 12-year record of such stations.

The results of this study are a general qualitative indication of long- and short-term records. They show, for example, that if we do not demand too great a degree of accuracy, the 50- and 100-year floods may be determined from the 40- or 50-year records which are commonly available.

It seems that the individual short-term station record is perhaps less reliable than we have generally considered it, but the individual long-term record is surprisingly dependable.

For less than the 100-year flood, a longer record than the period of the desired flood is necessary, for the result to be within 10 percent of the correct answer. This is increasingly true for the short-term floods. The correct answer for short-term floods probably would be 10 percent or less. In general, we should feel content if we are reasonably sure of predicting a given flood within 25 percent.

The figures shown cannot be expressed as percentages of the correct values and applied generally. The percentages are dependent on the slope of the individual frequency curve in question.

Using the theory of extreme values, theoretical curves can be derived showing, for 95 percent expectancy, the range of recurrence interval which might be assigned to an actual flood of known recurrence interval. This set of curves for the 10-year flood is shown as figure 26. (This is the same set of curves used in the homogeneity test.) On figure 26 are plotted the recurrence intervals corresponding to the actual 10-year flood (4,860 cfs), as determined from the 170 individual frequency curves of this study. Theoretically, 5 percent of these points, or 8.5 points should lie outside the limits; actually, 8 lie outside.

COMPOSITE FLOOD-FREQUENCY CURVES

The preceding section gives actual figures on the variability of results from different periods of record. Figures 20 to 23 show the same record graphically. Figure 20 illustrates the variability among individual short-term records. Many of the records are short, but in spite of their unreliability each one is valuable in adding to the accuracy of the basic relations to be established.

The Geological Survey makes predictions of frequency on regional flood-frequency studies where possible. For methods of regional analysis see Dalrymple (1950), Carter (1951), and other published reports. Evidence indicates that a flood-frequency graph that is based on the combined experience of a group of stations has firmer support than one drawn to fit the data at a single station. The process of combining the individual records tends to minimize the effect of the erratic samples.

In combining a group of station records, the individual records are not combined end-to-end to produce a long-term record. This procedure, (the "station-year method") is not considered justifiable in the use of discharge records. Combining ten 10-year records does not give a 100-year record but can provide a more dependable 10-year record, if records for homogeneous regions are being combined.

The data for this study afford an opportunity for a general check on these methods. Individual station records in the same general area show about as much variation as appears in the curves of figures 20, 21, 22, or 23. Therefore it was justifiable to treat them as though they were individual records in producing composite curves, and the results probably are comparable with those found in combining separate station records.

PERIODS COMBINED

Using procedures as outlined according to Dalrymple (1950) and Carter (1951), the short periods of record were combined as follows:

1. The first 10, second 10, and so forth of the 100 10-year periods were combined, giving a total of 10 composite graphs.

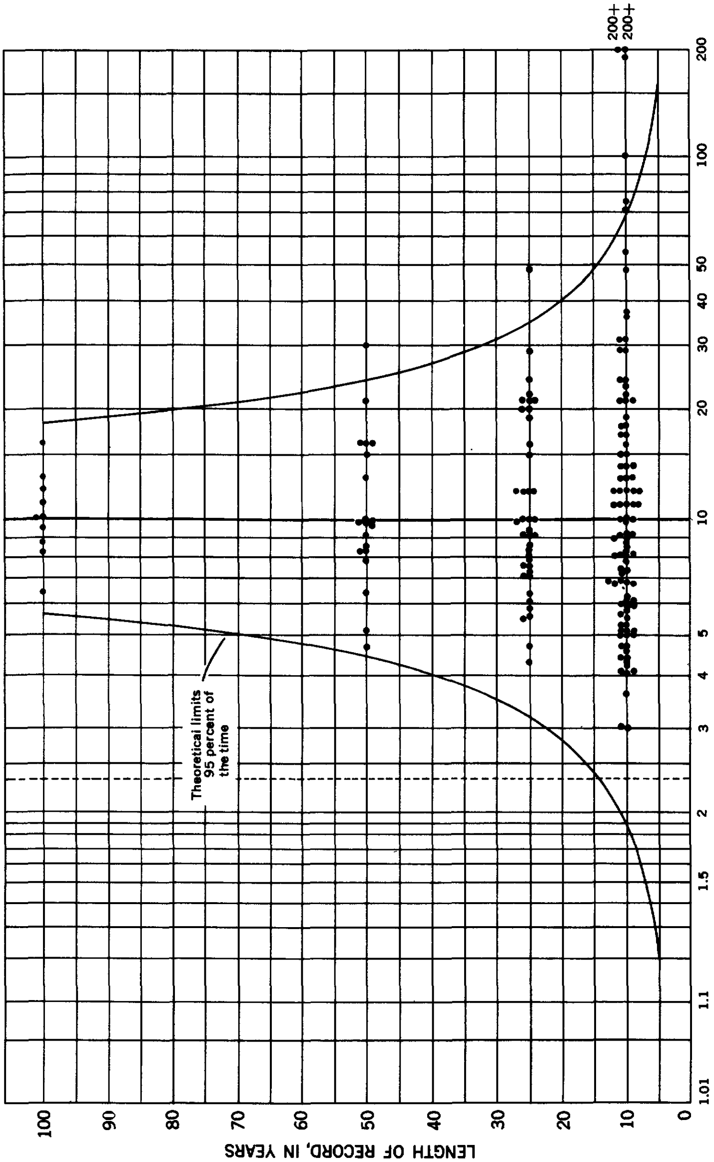

FIGURE 26.—Recurrence intervals of 10-year flood.

2. The first 10, second 10, and so forth, of the forty 25-year periods were combined, giving 4 composite graphs.
3. The first 10 and second 10 of the twenty 50-year periods were combined, giving 2 composite graphs.

For each single record, the separate items were expressed as ratios to the individual graphic means. Homogeneity tests were applied and in a few examples a record was discarded and not combined with the group. This practice was followed although the true mean was known and the separate parts all came from a single and therefore homogeneous record. The process was the same as is followed when combining actual station records.

The composite graphs derived in this manner are shown in figures 27 to 29 for combined 10-, 25-, and 50-year periods, respectively.

RESULTS FROM COMPOSITE GRAPHS

Analysis of these curves shows that the composites of the 10-, 25-, and 50-year periods define the 100-year flood within 22.7, 18.2 and 4.5 percent respectively. These are the outside limits represented by the curves—most of the values are much closer. These results cannot be applied quantitatively, however they are near the correct value. By means of any one of these combinations of 10-year periods, the 100-year flood could be predicted within 22.7 percent.

EFFECT OF NUMBER OF STATIONS COMBINED

An investigation was made of the accuracy of results as affected by the number of stations used to define a composite curve. Five 10-year periods were chosen at random, and a composite curve derived from them. Five more periods were chosen at random, and a composite curve based on these 10. The process was repeated up to the final combining of 20 records.

Results from the first five periods were so near the true values that there was not much opportunity for improvement with an increasing number of records. A second trial was therefore made repeating the entire process. For this trial, the values jumped erratically and showed no improvement with number—in fact, the results of combining 20 records were less accurate than for 5 records.

After a large number of records have been combined (possibly 50 or 100), there should be a definite trend toward the correct answer. Apparently with a small number of records, and both 5 and 20 are small numbers statistically in this sense, the results of chance may be very erratic. There may be available 5 records whose composite answer is near the true values, or there may be 20 records whose composite result is less accurate. However, the results of using composite frequency curves seems to be good enough even with a small number of stations available. Using medians and combining

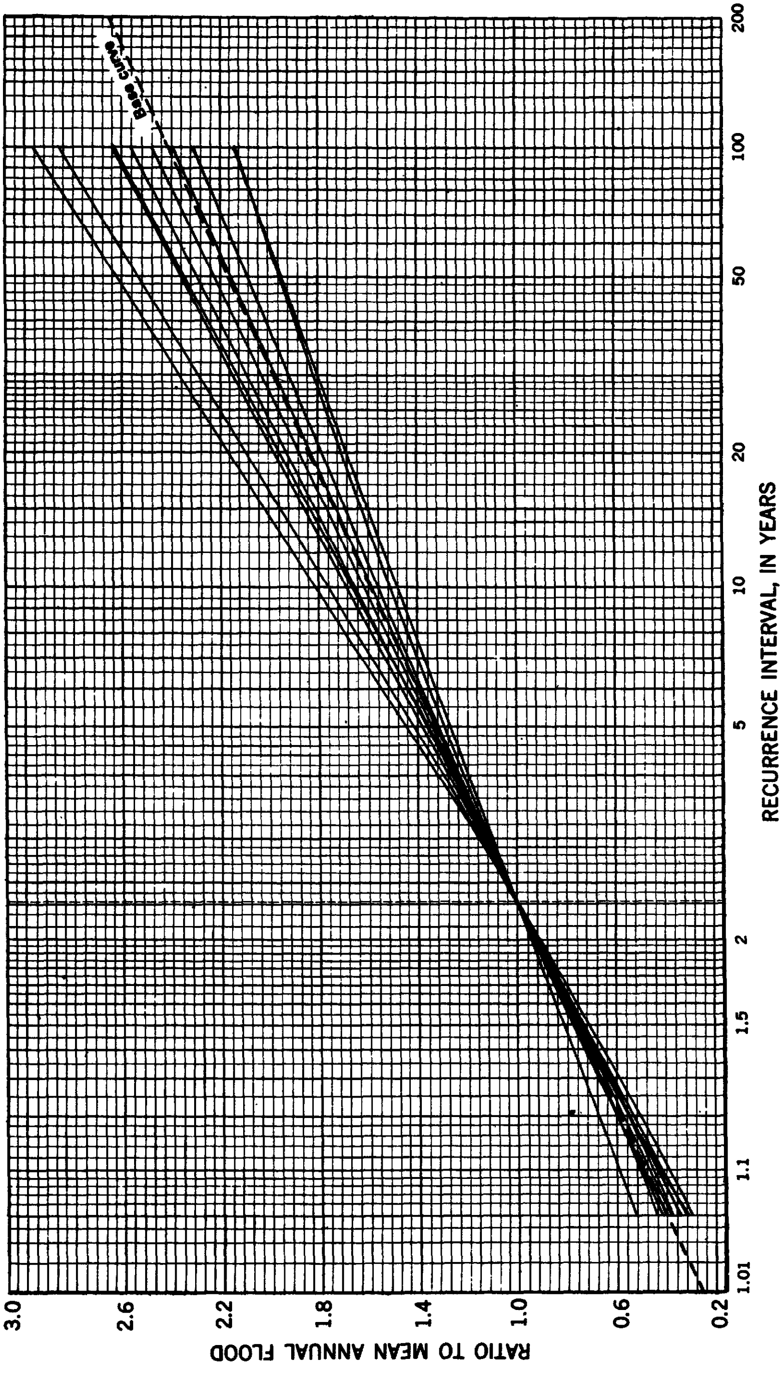

FIGURE 27.—Composite frequency curves based on ten 10-year periods.

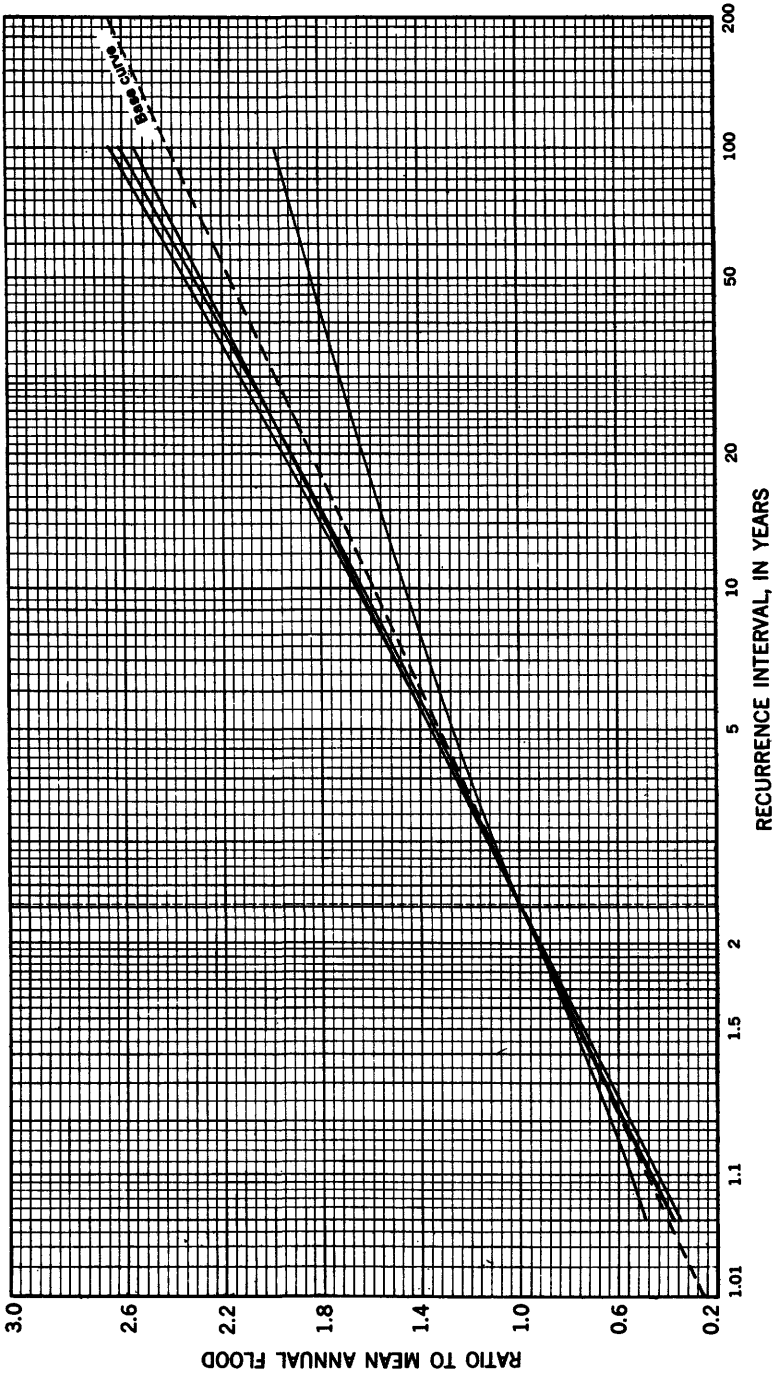

FIGURE 28.—Composite frequency curves based on ten 25-year periods.

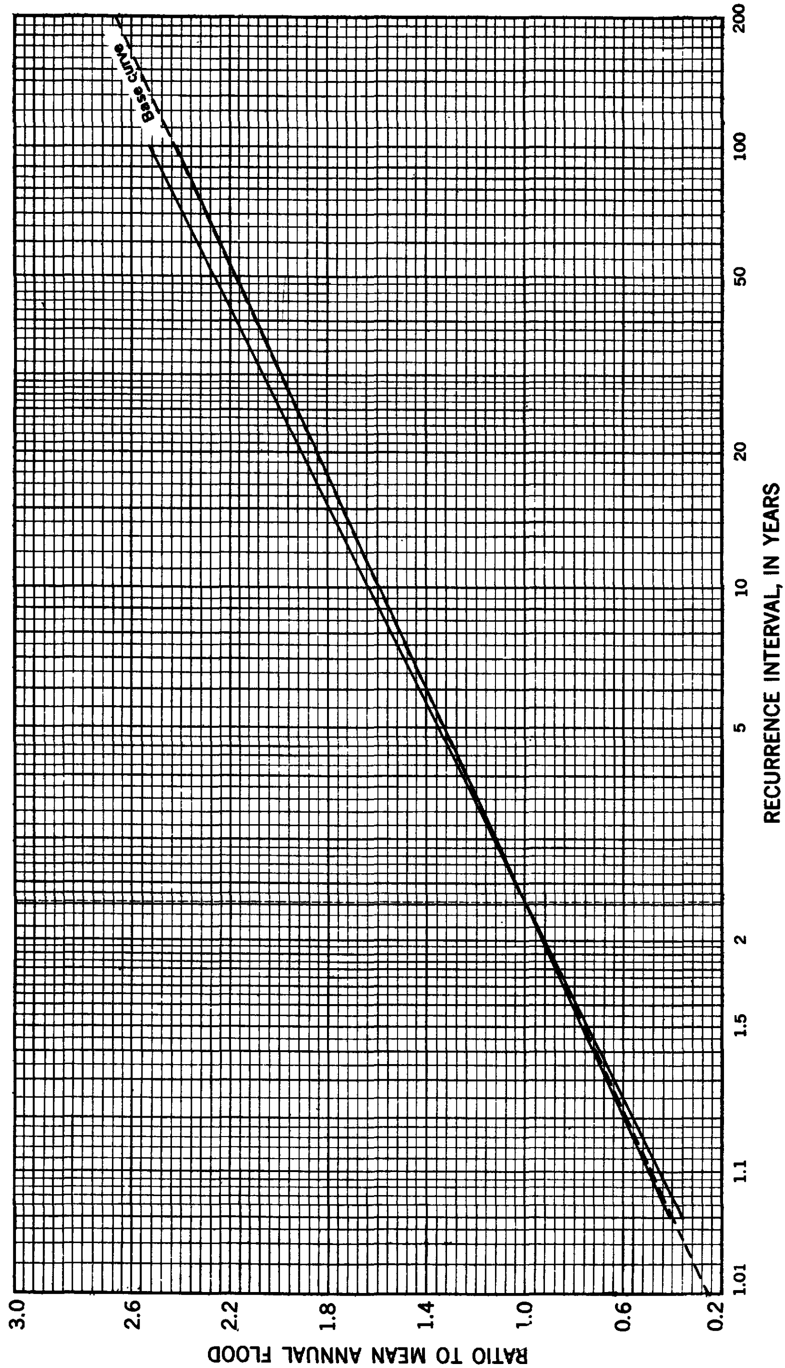

FIGURE 29.—Composite frequency curves based on ten 50-year periods.

from five to twenty 10-year records, the 10-year flood could be predicted with 5.7 percent, the 25-year flood within 7.3 percent, the 50-year flood within 7.8 percent, and the 100-year flood within 8.7 percent.

Statistical theory gives some indication of what to expect from increasing samples. The results of composite curves may be considered as mean values and probably the results have somewhat the same characteristics as means. Statistics state that the standard deviation of the mean of samples drawn from a single population varies inversely as the square root of the sample size. The standard deviation is a measure of the expected departure from the true value, and samples of 20 should have one-half the standard deviations of samples of 5, on a long-time average.

MEDIAN VERSUS AVERAGE RATIOS

In combining the dimensionless flood ratios in the process of obtaining a composite graph, the median value of the ratios for any specific order number has ordinarily been used (Dalrymple 1950, p. 15, col. 9). There is some question as to whether it might be preferable to use the arithmetic average instead of the median.

A comparison of the two has been made for all the composite graphs developed in this study. Because the correct values of any frequency of flood are known, the departures from the correct values are an indication of the validity of any procedure.

The ranges in value of the 10-, 25-, 50-year floods, based on records of various lengths, show that use of the average rather than the median flood ratios practically does away with the apparent inconsistency, previously mentioned, shown by 10- and 25-year records. However, both the minimum and maximum values are increased.

Studies were made of the values of 10-, 25-, 50-, and 100-year floods based on the 2 trials in which the number of combined stations was increased from 5 to 20. The values based on averages were generally less accurate than those based on medians. Also, those based on averages were predominantly too high.

On the basis of these two sets of comparisons, the medians have an advantage on the averages, and it seems best to continue the use of medians as recommended. The use of medians rather than averages agrees with the use of graphical rather than arithmetic means. Both procedures tend to nullify the effect of abnormally large floods in short-term records.

THEORETICAL DISTRIBUTION OF COMPOSITE CURVES

The results of using composite curves were so favorable that a theoretical study was made to confirm these specific results.

The following discussion refers to figure 30. Lines *A* and *F* represent the extreme positions of a group of frequency curves based on the same period of record such as the group of curves in figure 20.

Lines *B*, *C*, *D*, and *E* divide the total range into 5 areas 1, 2, 3, 4, 5), each containing 20 percent of the entire number of curves. Because the curves do not necessarily parallel these lines, but may cross them, the lines, more precisely, have such a position that for any recurrence interval one-fifth of the curve intersections would lie within each area.

All points within each of the areas 1, 2, 3, 4, and 5 have been represented by average numerical values of 1.0, 2.0, 3.0, 4.0, and 5.0. The positions of the averages of several points are represented by the numerical average of their individual values. Where the analysis applies equally at any recurrence interval, applies to the composite frequency curve.

Consider a curve in area 1 and a curve in area 5; if they are combined they will have a position somewhere in area 3. This is repre-

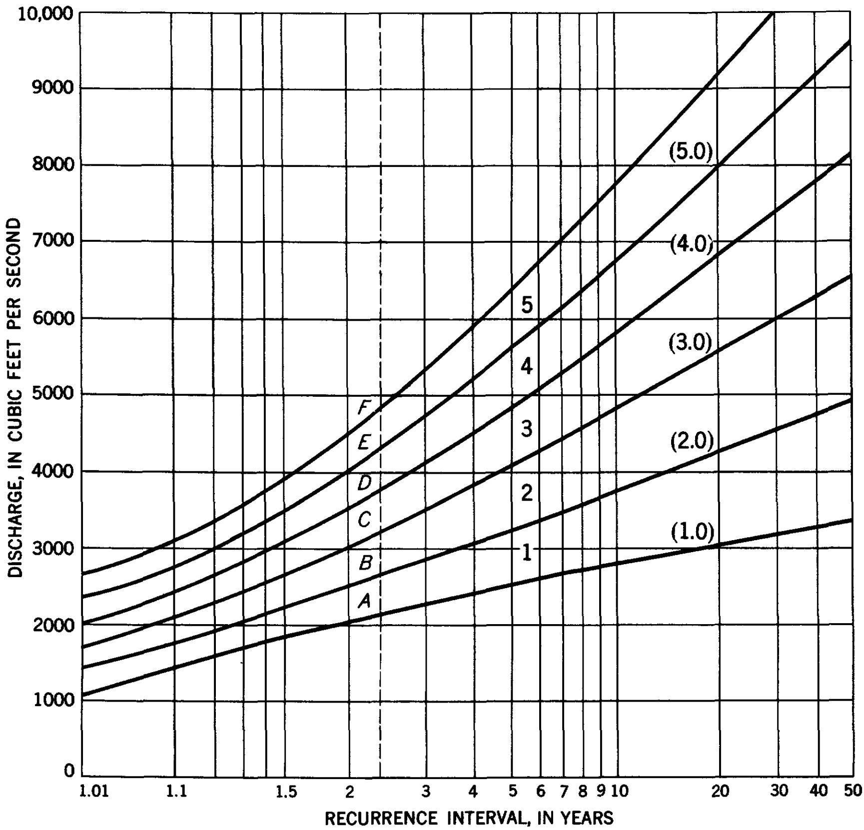

FIGURE 30.—Theoeretical distribution of composite curves.

sented numerically by the average of their numerical values, 1.0 and 5.0, is 3.0. The distribution is slightly unbalanced, but in actual practice, medians are used rather than averages, and the effect of medians is to lower the result somewhat.

The effect of combining any five curves can be represented by the average of their numerical values. If all 5 curves were in area 1, the numerical value of their average would be 1.0; if all 5 curves were in area 5, their average would be 5.0. There are 3,125 (that is 5^5) possible variations within the 5 areas of 5 curves chosen at random. These may be considered as all the possible combinations of 5 numbers, each of which may have the value 1, 2, 3, 4, or 5. The average of each combination is then computed. All averages lying between 1.0 and 1.5 are considered as lying in area 1, averages between 1.5 and 2.5 as lying in area 2, and so forth. The distribution of all the 3,125 means is shown below in percentage.

The number of stations combined in regional frequency studies generally is 10 or more. If 10 records are combined, there are 9,765,625 (that is, 5^{10}) possible variations. By statistical theory, (using the binomial theorem), the variation of the means has been computed and the percentage of distribution is as follows:

Percentage of distribution

Area	5 records combined	10 records combined
1	0. 7	0. 04
2	21. 3	13. 1
3	56. 0	73. 7
4	21. 3	13. 1
5	. 7	. 04
Total	100. 0	100. 0

These results indicate that the composites of either 5 or 10 stations have only a negligible chance of lying in either areas 1 or 5. With 5 records combined, chances are better than 50–50 that the composite curve will lie within the narrow band of area 3, and with 10 combined records, chances are 3 out of 4 that it will do so.

CONCLUSIONS

1. A confirmation by actual computation is obtained of the theoretically derived value of 2.33 years as the recurrence interval for the mean annual flood, for a straight-line distribution on a Gumbel graph.

2. A general confirmation is obtained of the method of drawing frequency graphs as visually best-fitting curves.

3. The graphical mean gives good results in general and is to be preferred to the arithmetic mean since it avoids large errors due to the occurrence of large floods within a short period.

4. The reliability of the mean annual flood increases with length of record. Short periods of record do not define it closely enough for some purposes. A method is available whereby a "working range" can be computed and used in correlation studies.

5. The reliability of flood magnitudes other than the mean also increases with length of record. Short periods of record (up to about 25 years) cannot define satisfactorily even short-term floods. Long-term records (40 to 50 years or longer) can define flood magnitudes up to the length of record or longer, depending upon the accuracy required. The individual short-term record is perhaps less reliable than generally considered, but the individual long-term record is surprisingly dependable.

6. Accuracy within 10 percent from individual station records is rarely attainable. Accuracy within 25 percent seems attainable with a fair length of record.

7. The results of combining even short-term records, similar to procedures in regional flood studies, are good.

8. The number of station records generally combined is always statistically small, and, a combination of 5 records may give as good results as a combination of 20 records. However, even a combination of five records can give reliable results.

9. Accuracy of 10 percent seems indicated from composite curves for frequencies ordinarily desired (50 years or less) based on records of the number and length generally available.

10. Theoretical consideration of the results of combining frequency curves shows a high degree of probability for favorable results.

SELECTED REFERENCES

American Society Civil Engineers, 1951, Review of flood frequency methods: Proc. Separate 110; 1953, Trans., v. 118, p. 1220–1230.

Anderson, H. W., 1949, Flood frequencies and sedimentation from forest watersheds: Am. Geophys. Union Trans., v. 30, p. 567–584.

Arkin, Herbert, and Colton, Raymond, 1939, An outline of statistical methods: New York, Barnes & Noble, 220 p.

Beard, L. R., 1943, Statistical analysis in hydrology: Am. Soc. Civil Engineers Trans., v. 108, p. 1110–1160.

——— 1952, Statistical methods in hydrology: U.S. Army, Corps of Engineers, Civil Works Engineer Bull. 52–24.

——— 1954, Estimation of flood probabilities: Am. Soc. Civil Engineers Proc. Separate 438.

——— 1956, Statistical evaluation of runoff volume frequencies: Internat. Assoc. Hydrology (de l' U.G.G.I.), Symposia Darcy, pub. no. 42, p. 171–183.

Benson, M. A., 1950, Use of historical data in flood-frequency analysis: Am. Geophys. Union Trans., v. 31, p. 419–424.

Benson, M. A., 1952, Characteristics of frequency curves based on a theoretical 1,000-year record: U.S. Geol. Survey open-file report.

Bernard, M. M., 1936, Determination of flood flow by unit-hydrograph method, *in* Jarvis, C. S., and others, Floods in the United States, magnitude and frequency: U.S. Geol. Survey Water-Supply Paper 771, p. 451–461.

Berwick, V. K., 1958, Floods in eastern Montana, magnitude and frequency: U.S. Geol. Survey open-file report.

Bigwood, B. L., and Thomas, M. P., 1955, A flood-flow formula for Connecticut: U.S. Geol. Survey Circ. 365, 16 p.

Bodhaine, G. L., and Robinson, W. H., 1952, Floods in western Washington, frequency and magnitude in relation to drainage-basin characteristics: U.S. Geol. Survey Circ. 191, 124 p.

California Department of Public Works, 1923, Flow of California streams: Div. Engineering and Irrigation, Bull. 5, chap. 5.

Carter, R. W., 1951, Floods in Georgia, frequency and magnitude: U.S. Geol. Survey Circ. 100, 127 p.

Chow, Ven Te, 1950, Discussion of Annual floods and the partial duration flood series, by W. B. Langbein: Am. Geophys. Union Trans., v. 30, p. 939–941.

——— 1951, A general formula for hydrologic frequency analysis: Am. Geophys. Union Trans., v. 32, p. 231–237.

——— 1953, Frequency analysis of hydrologic data with special application to rainfall intensities: Illinois Univ. Eng. Expt. Sta. Bull. 414.

——— 1954, The log-probability law and its engineering applications: Am. Soc. Civil Engineers Proc. Separate 536.

——— 1956, Hydrologic studies of floods in the United States: Internat. Assoc. Hydrology (de l' U.G.G.I.), Symposia Darcy, pub. no. 42, p. 134–170.

Cragwall, J. S., Jr., 1952, Floods in Louisiana, magnitude and frequency: Louisiana Dept. Highways, 281 p.

Creager, W. P., Justin, J. D., and Hinds, J., 1944, Engineering for dams: New York, John Wiley & Sons, v. 1, 245 p.

Cross, W. P., 1946, Floods in Ohio, magnitude and frequency: Ohio Dept. Public Works, Water Resources Board Bull. 7, 154 p.

Dalrymple, Tate, 1950, Regional flood frequency: Highway Research Board, Research Rept. no. 11–B, p. 4–20.

——— 1957, Flood frequency relations for gaged and ungaged streams: Internat. Assoc. Hydrology (de l' U.G.G.I.), Gen. Assembly Toronto, pub. no. 45, p. 268–279.

Ezekiel, Mordecai, 1945, Methods of correlation analysis: New York, John Wiley & Sons, 522 p.

Fisher, R. A., 1946, Statistical methods for research workers: London, Oliver & Boyd, 354 p.

Fisher, R. A., and Tippett, L. H. C., 1928, Limiting forms of the frequency distribution of the largest or smallest member of a sample: Proc. Cambridge Philos. Soc., v. 24, p. 180–190.

Foster, E. E., 1948, Rainfall and runoff: New York, Macmillan Co., 487 p.

Foster, H. A., 1924, Theoretical frequency curves and their application to engineering problems: Am. Soc. Civil Engineers Trans., v. 87, p. 142–203.

Furness, L. W., 1955, Floods in Nebraska, magnitude and frequency: Nebraska Dept. Roads and Irrigation, 103 p.

Grover, N. C., 1937, The Floods of March, 1936, part 1, New England rivers: U.S. Geol. Survey Water-Supply Paper 798, 466 p.

Grover, N. C., 1937, The Floods of March, 1936, part 2, Hudson River to Susquehanna River region: U.S. Geol. Survey Water-Supply Paper 799, 380 p.

——— 1937, The Floods of March, 1936, part 3, Potomac, James, and upper Ohio Rivers: U.S. Geol. Survey Water-Supply Paper 800, 351 p.

Gumbel, E. J., 1941, The return period of flood flows: Annals Math. Statistics, v. 12, no. 2, p. 163–190.

——— 1942, Statistical control-curves for flood-discharges: Am. Geophys. Union Trans., pt. 2, p. 489–509.

——— 1945a, Floods estimated by probability method: Eng. News-Rec., v. 134, no. 24, p. 833–837.

——— 1945b, Simplified plotting of statistical observations: Am. Geophys. Union Trans., v. 26, p. 69–82.

——— 1948, The statistical forecast of floods: Ohio Water Resources Board pamphlet.

——— 1952, Discussion *of* A general formula for hydrologic frequence analysis, by Ven Te Chow: Am. Geophys. Union Trans., v. 33, p. 277–282.

——— 1954, Statistical theory of extreme values and some practical applications: U.S. Bur. Standards Appl. Mathematics, ser. 33, 51 p.

——— 1958, Statistics of extremes: New York, Columbia Univ. Press, 371 p.

Gumbel, E. J., and Von Schelling, H., 1950, The distribution of the number of exceedances: Annals Math. Statistics, v. 21, no. 2, p. 247–262.

Hazen, Allen, 1930, Flood flows: New York, John Wiley & Sons, 199 p.

Hoel, P. G., 1947, Introduction to mathematical statistics: New York, John Wiley & Sons, 258 p.

Horton, R. E., 1936, Hydrologic conditions as affecting the results of the application of methods of frequency analysis to flood records, *in* Jarvis, C. S., and others, Floods in the United States, magnitude and frequency: U.S. Geol. Survey Water-Supply Paper 771, p. 433–450.

Hoyt, W. G., and others, 1936, Studies of relations of rainfall and run-off in the United States: U.S. Geol. Survey Water-Supply Paper 772, 301 p.

Jarvis, C. S., and others, 1936, Floods in the United States, magnitude and frequency: U.S. Geol. Survey Water-Supply Paper 771, 497 p.

Johnstone, Don, and Cross, W. P., 1949, Elements of applied hydrology: New York, Ronald Press, 276 p.

Kimball, B. F., 1946, Assignment of frequencies to a completely ordered set of sample data: Am. Geophys. Union Trans., v. 27, p. 843–846.

Kinnison, H. B., and Colby, B. R., 1945, Flood formulas based on drainage basin characteristics: Am. Soc. Civil Engineers Trans., v. 110, p. 849–904.

Langbein, W. B., 1949, Annual floods and the partial-duration series: Am. Geophys. Union Trans., v. 30, p. 879–881.

Langbein, W. B., and others, 1947, Topographic characteristics of drainage basins: U.S. Geol. Survey Water-Supply Paper 968–C, p. 125–157.

Lieblein, Julius, 1954, A new method of analysing extreme-value data: U.S. Bur. Standards, Nat. Advisory Comm. Aeronautics Tech. Note 3053, 88 p.

Linsley, R. K., Kohler, M. A., Paulhus, J. L. H., 1949, Applied hydrology: New York, McGraw-Hill Book Co., 689 p.

McCabe, J. A., 1957, Floods in North and South Dakota, frequency and magnitude: U.S. Geol. Survey open-file report.

——— 1958, Floods in Kentucky, magnitude and frequency: U.S. Geol. Survey open-file report.

Mitchell, W. D., 1954, Floods in Illinois, magnitude and frequency: Illinois Dept. Public Works and Buildings, Div. Waterways, 386 p.

Noecker, Max, 1952, Floods in Youghiogheny and Kiskiminetas River basins, Pennsylvania and Maryland, frequency and magnitude: U.S. Geol. Survey Circ. 204, 22 p.

Peirce, L. B., 1954, Floods in Alabama, magnitude and frequency: U.S. Geol. Survey Circ. 342, 105 p.

Potter, W. D., 1949, Effect of rainfall on magnitude and frequency of peak rates of surface runoff: Am. Geophys. Union Trans., v. 30, p. 735–751.

Powell, R. W., 1943, A simple method of estimating flood frequencies: Civil Eng. v. 13, no. 2, p. 105–106.

Pride, R. W., 1957, Flood frequency relations for Florida: U.S. Geol. Survey open-file report.

Prior, C. H., 1949, Magnitude and frequency of floods in Minnesota: Minnesota Dept. Conserv., Div. Waters, Bull. 1, 128 p.

Rantz, S. E., and Riggs, H. C., 1949, Magnitude and frequency of floods in the Columbia River Basin, *in* Floods of May–June 1948 in Columbia River Basin: U.S. Geol. Survey Water-Supply Paper 1080, p. 317–469.

Rider, R., 1939, An introduction to modern statistical methods: New York, John Wiley & Sons, 220 p.

Riggs, H. C., 1955, Floods in North Carolina, magnitude and frequency: U.S. Geol. Survey open-file report, 59 p.

Saville, Thorndike, 1936, A study of methods of estimating flood flows applied to the Tennessee River, *in* Jarvis, C. S., and others, Floods in the United States, magnitude and frequency: U.S. Geol. Survey Water-Supply Paper 771, p. 398–420.

Schwob, H. H., 1953, Iowa floods, magnitude and frequency: Iowa Highway Comm., Iowa Highway Research Board Bull. 1, 171 p.

Searcy, J. K., 1955, Floods in Missouri, magnitude and frequency: U.S. Geol. Survey Circ. 370, 126 p.

Siegel, S., 1956, Non parametric statistics: New York, McGraw-Hill Book Co., 312 p.

Slade, J. J., Jr., 1936, The reliability of statistical methods in the determination of flood frequencies, *in* Jarvis, C. S., and others, Floods in the United States, magnitude and frequency: U.S. Geol. Survey Water-Supply Paper 771, p. 421–432.

Smith, Winchell, and Heckler, W. L., 1955, Compilation of flood data in Arizona, 1862–1953: U.S. Geol. Survey open-file report.

Tice, R. H., 1958, Delaware River basin flood frequency: U.S. Geol. Survey open-file report, 10 p.

Velz, C. J., [1950–51], Graphical approach to statistics: reprinted 1952, Chicago, Water & Sewage Works, 30 p.

Woolley, R. R., 1946, Cloudburst floods in Utah, 1850–1938: U.S. Geol. Survey Water-Supply Paper 994, 128 p.

INDEX

Page

Acknowledgments 1
Adjustment from short to long period 40, 42
Annual flood, definition 5
Annual-flood series 5, 6, 8
Annual reports, preparation 11

Bankfull stage 17
Bar graph 34, 45
Base discharge, selection 11, 12, 13
Base period, adjustment 34, 45
selection 32
Basic-frequency curve 26
Basin lag 30
Beard method 16
Big Piney Run near Salisbury, Pa 35, 37

"California" method 15, 16, 50
Channel storage, effect on mean annual flood 41, 43
Combining of short periods 65
Combining records within a region 25, 73
Composite curves, theoretical distribution 71
Composite factors 30
Correlation with physical factors 43
Cyclic changes in weather 32

Data, listing 8–13
Degrees of freedom 31
Dimensionless frequency curves 26, 27
Discharge-frequency curve 5
Discharge records, length 3, 18, 20
Drought years, effect on discharge 43

Envelope curve 21
Estimation of years of no record 35
Examples of regional reports 47
Exceedences, distribution 49
Extrapolation 20, 21

Fitting curves to data, graphically 3, 20, 42, 43, 54, 56, 73
mathematically 3, 20, 37, 56
Flexible skew-distribution curve 2
Flood-control storage 8
Flood-frequency curves, for short-term periods 21, 24, 25, 40, 51, 54, 64, 74
types 5, 7, 8
Flood-frequency method, application 2, 3, 5, 20, 21, 25
Flood-frequency relation, regional 24, 25, 26, 27, 39, 43, 45–46, 61, 65, 73
Flood-frequency, relation to drainage-basin characteristics 4, 26, 29–31, 41, 43, 61
Flood-frequency reports, Statewide or regional 1, 3, 4, 24, 47
Flood insurance 2
Flood magnitudes, reliability 63
Flood potential of a region 64
Flood stages, significance 10
Flood zoning 2
Floods, considered in structure design 2
magnitude and frequency 2, 3
Frequency curve, definition 45

Page

Gaging-station data, in water-supply papers 11
Geological Survey method 3, 4, 6, 16, 49
Gumbel method 3, 15, 16, 28, 49, 50, 73

Handbook for Hydrologists 2
Hazen method 3, 15, 51
Historical flood data 8, 13, 17, 18, 37
Homogeneity test 27–29, 38, 39, 45, 65, 67
Hurricane conditions, effect on flood peaks 15
Hydrologic areas 31

Ice-affected streams 7, 10
Interpolation 20

Licking River at Toboso, Ohio 13, 14
Logarithmic discharge scales 19, 21, 37
Logarithmic skew-frequency curve 3

Mean annual flood, definition 4, 26, 29, 31, 37, 54
estimation 41, 55–63
Mean annual rainfall 30
Mean annual runoff 30
Median flood ratio 39, 45, 46
Meteorologic factors 30
Mississippi River, selection of stations on 31
special treatment for 46
Multiple correlation techniques 43

Newspaper sources for historical information 13
Nomographs 43
Nueces River below Uvalde, Tex 13
Number of stations, effect on results 67

Olentangy River near Delaware, Ohio 44
Order numbers, adjustment 17, 18
computation 36

Partial-duration series 5, 6, 13, 16, 19, 21, 43, 44
Physiographic factors 29
Plotting paper, preparation 19, 21
Plotting positions, computation 5, 15, 17, 18, 49–51
Preliminary frequency curve 37, 39; fig. 11

Rainfall-intensity frequencies 25
Range line 42, 61
Rectangular discharge scales 19, 21
Recurrence interval, computation 4, 15, 17, 18, 19, 20, 28, 29, 31, 37, 38, 45, 47, 49, 54, 63.
definition 5, 6, 15, 24, 72
Red River in Louisiana 47
Regional frequency curve 27, 39, 40, 41, 43, 46, 65
Relation between annual-flood and partial-duration series 6
Reliability of mean annual-flood values 61
Rules for publishing peaks 11

Sampling errors 2, 3, 24, 25, 37, 56
San Antonio River at San Antonio, Tex 13
Single-station curve 24
Skewness of the data 3, 15
Snowmelt, effect on peak discharge 12, 15
Stage-discharge relation 7

Page

Stage-frequency curve 5
Standard deviation, definition 71
Station-year method of analysis 25
Straight-line graphs 55, 56
Stream channels, shifting 7
Streams, manmade regulation 11
Susquehanna River at Harrisburg, Pa 13, 18
Suwanee River at White Springs, Fla 10

Theoretical 1,000-year record 24, 26, 27, 28, 51, 53
Time graduation, equation 19
Total usable storage capacity 31

Page

Use of short-term records 43

Volume-frequency curve 5

Working range of mean annual floods 61
Work sheets, listing of data on 34
Works of man, effect on floodflow 4, 31

Youghiogheny and Kiskiminetas River basins,
flood-frequency curve defined 45
stations used in analysis 32, 33

O

U.S. GOVERNMENT PRINTING OFFICE : 1960 O - 540570