REGIONAL FLOOD FREQUENCY

Tate Dalrymple
U. S. Geological Survey

SYNOPSIS

The practice of the United States Geological Survey in determining the magnitude and frequency of floods is presented. The flood frequency analyses described herein are based on the discharge records that have been obtained as a part of their nation-wide stream-gaging program. Gaging station records are obtained at specific locations and seldom are available at the site of a proposed structure. For maximum usefulness these records must be made applicable to ungaged areas.

A method of computing flood frequencies at a gaging station is first described. The annual flood method rather than the partial duration series method is recommended, due to simplicity. A definite relationship between the two has been established and values on a partial duration series basis, useful in the design of overflow type of bridge approach, can be obtained from the study of annual floods.

The next part of the paper shows how gaging station records may be combined to give a frequency relationship applicable for a region. This method provides a means for estimating flood frequencies on ungaged streams. Two curves usually result from the regional study, one showing the ratio of the mean annual flood to a flood of any selected frequency and the other showing the relationship between the mean annual flood and some function of the drainage area, usually the size. From these two curves the magnitude and frequency of a flood at any location on any stream in the region may be determined. Due to lack of basic data, results are generally restricted to areas larger than 100 sq. mi. and for frequencies less than 50 years.

An example is worked out showing application of the method to the Maumee River Basin in Ohio, Indiana and Michigan. A short bibliography relating to both statistics and floods completes the paper.

The purpose of this paper is to present a method of analyzing streamflow data to obtain information relative to the magnitude and frequency of flood discharges. Studies of flood frequencies have been found especially helpful in problems involving economic considerations such as design of bridge clearances, channel capacities and roadbed levels, where costs must be balanced against flood damage or liabilities arising from failure and interruption of services. Drainage structures are seldom designed capable of passing the maximum flood that may occur as it is not economically sound to provide for such unusual occurrences. A flood less than the maximum possible usually will be selected as the basis for design, and some consideration must be given to the frequency with which this flood will recur.

The subject of flood frequencies has attracted many investigators and much profit has been derived from their study, but the viewpoints and theories expressed have not always been consistent. There is no uniformity of opinion today as to which is the best method, and probably one is as good as the other; although results may differ somewhat this is not serious provided consideration is given to the particular method used.

The method of computing flood frequencies that is presented in this paper reflects the latest developments based on a continuing study of the subject by engineers of the Water Resources Division of the United States Geological Survey. The method has been revised several times in the past few years and probably will be again in the future. There is no static level for such a rapidly developing subject.

The discharge records collected by the Geological Survey, and other private, State and Federal agencies, are the data upon which flood frequency studies are made. These records are obtained at about 6,000 places in the United States and, where of sufficient length, furnish an excellent base for deriving a flood frequency curve. Such a curve, however, applies only to that one particular place; generally the information is wanted at an ungaged spot. Thus there are two parts to the problem: (1) computation of flood frequency at a point, as a gaging station, and (2) transferring this point data to other places or adapting it to apply over a basin or a region.

FLOOD FREQUENCY AT A GAGING STATION

The first step in beginning a flood compilation is to select the gaging stations to be included. Gaging stations to be considered for this kind of study should have at least 10 years of record, although records as short as 5 years have been used when no others were available on any nearby stream. Storage or other artificial factors which would tend to modify flood discharges significantly should be a minimum. Do not include canals, ditches, and drains in which discharges are subject to substantial control by man.

Peak stages and corresponding discharges both should be listed in chronologic order. Momentary peaks only should be listed, although mean daily discharges have been used extensively in the past. For streams with loop ratings, or those subject to rate of change or backwater effects, peak stages are not concurrent with peak discharges and they should be listed independently.

The next step is to review the station history, especially the stage-discharge rating. The determination of the proper rating and its use to obtain peak discharges requires careful study and considered judgment. As this review is perhaps the most important and valuable part of a flood-frequency investigation, it should be performed by persons experienced in field and office stream-gaging practice. The desired goal is to obtain a set of peak discharges that are consistent among themselves and as accurate as all available data permit. A common source of inconsistency arises from changes in rating curves on the basis of recent discharge measurements and neglect to revise past records. No way has been found to assure that all users will obtain revised figures of previously published records.

Where it is necessary to compute frequencies of stage occurrence careful thought must be given to the nature of the stage-discharge relation. If the stage-discharge relation has remained essentially stable throughout the period of record, then frequencies can either be computed directly from stage or preferably can be computed from discharge first, then transferred to stage by means of the stage-discharge relation. In the case of sandy channels there may be very frequent shifting of the bed, and a stage-frequency curve may be valid only for a definite period of time or, in extreme cases, may be of little value. Ice-affected streams involve further problems yet unsolved; when 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.

Kinds of Flood Series - There are two kinds of flood series to be considered, annual floods and a partial duration series, or floods above a base.

An annual flood is defined as the highest momentary peak discharge in a water year. Only the greatest flood in each year is used. An objection most frequently encountered with respect to the use of annual floods is that it uses only one flood in each year. Infrequently, the second highest flood in a given year, which the above rule omits, may outrank many annual floods.

This objection is met by listing all floods above a selected base without regard to number within any given period. Such a list is called a partial duration series. The base is generally selected as equal to the lowest annual flood so that at least one flood in each year is included. In a long record, however, the base is usually raised so that on the average only three or four floods a year are included. The only other criterion to follow in the selection of the floods is that each peak be individual, i. e., be separated by substantial recession in stage and discharge. An objection to the use of the partial duration series is that the floods listed may not be fully independent events, i. e., one flood sets the stage for another.

the average interval in which a flood of given size will recur as an annual maximum. In the partial duration series, this is the average interval between floods of a given size regardless of their relationship to the year or any other period of time. This distinction remains, even though for large floods the recurrence intervals are closely the same on both scales. The two methods give essentially identical results for intervals greater than about

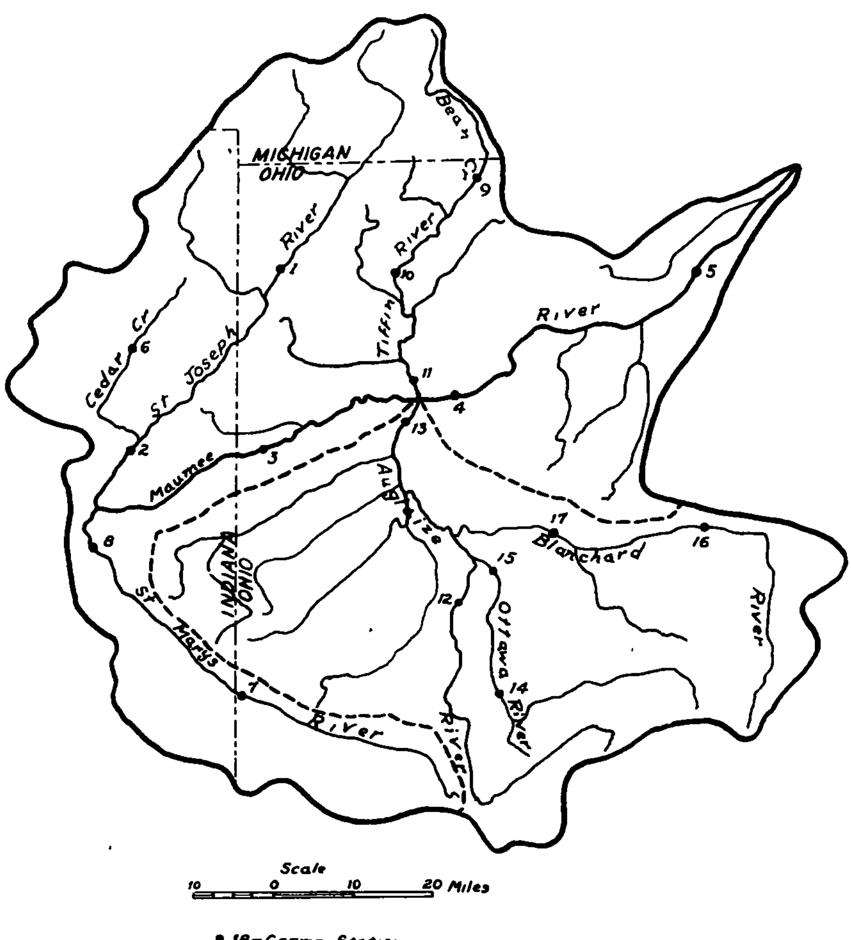

Figure 1. Maumee River Basin

There is an important distinction in meaning as between the recurrence intervals of annual floods and the partial duration series floods. In the annual flood series the recurrence interval is

ten years. As most designs are for intervals greater than this, it is apparent that, from a practical standpoint, either method is satisfactory, although the simplicity of the annual flood method makes it attractive.

From statistical principals there is a definite relationship between the values in the two series. The following tabulation shows comparative values of recurrence intervals by the two methods:

Recurrence Intervals in Years

Annual Floods	Partial Duration Series
1.10	0.41
1.25	.62
1.50	.91
1.75	1.18
2.00	1.45
2.5	2.0
5.0	4.6
10.0	9.5
15.0	14.5
20.0	20.0
100	100

Plotting Positions - Having the discharges listed, it is customary to number them in order of magnitude, numbering the largest 1. Then there is known (1) the relative distribution of floods in (2) a given period of years. The next step is to fit a time scale to the data. Published ideas on this subject are quite diverse, largely because people differ as to the proper method of treating small samples.

The formula adopted by the Geological Survey is simple to compute, is applicable both to annual flood data and the partial duration series, and gives results acceptably in conformance with some of the latest theories. Recurrence intervals for both annual floods and the partial duration series are computed from the formula (N + 1)/M, where N equals number of years of record and M equals relative magnitude of the event beginning with the highest as 1.

Annual Flood Peaks - List the highest observed peak in each water year in chronological order. Only complete years of stream-flow records can be included but historical flood data can often be used to advantage, particularly to assist in defining the upper end of the frequency curve. The peaks should be numbered in order of magnitude beginning with 1 for the highest and recurrence intervals, in years, computed by the formula (N + 1)/M.

Annual floods are now plotted on a special form developed by Powell[1] for analysis of flood frequencies by the Gumbel[1] method. The discharges are plotted to a linear scale as ordinate: the abscissa (scale of recurrence intervals) is specially graduated according to the theory of largest values.

For the general purpose of flood-frequency graphs the kind of graduations on the paper is of no great importance. However, it is desirable to have uniformity, and if a choice is to be made, the chart based on the theory of largest values has much to offer as flood discharges plotted on this chart approximate a straight-line graph.

Partial Duration Series - List all peaks whose discharge exceeds a chosen base discharge, regardless of the number of peaks occurring in a year. The number of peaks will be about three to four per year if the base is chosen from the list of annual flood peaks as a discharge (rounded upward) whose recurrence interval is 1.15 years.

List only the highest peak of two or more occurring within 48 hr. of each other, unless it is evident that the peaks in that period are independent, as may occur at times on flashy streams. The peaks can be considered independent if ground-water flow is reached between them. Do not list 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. For periods of diurnal peaks caused by snow melt list only the highest occurring during each distinct period of melting regardless of the fact that other peaks may fulfill the preceding requirements.

The peaks should be arranged in order of magnitude and assigned numbers corresponding to their position in the array beginning with the highest as 1. The next step is to compute recurrence intervals for this class of floods

[1]See list of references at the end of this paper.

by the formula (N + 1)/M where N is number of years of record and M is order of magnitude.

The data should be plotted on semi-log graph paper, using the linear scale (ordinate) for the discharge data, and the logarithmic scale (abscissa) for recurrence intervals.

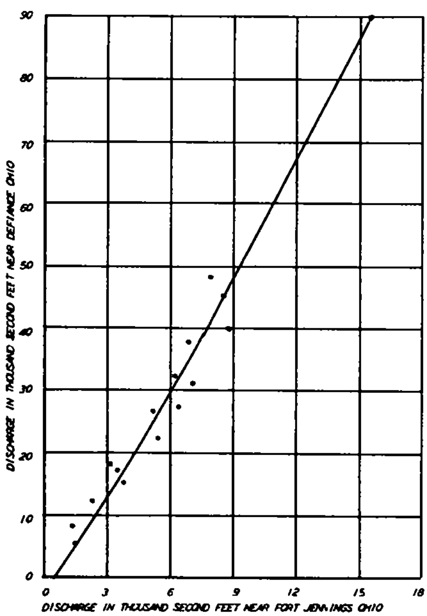

Figure 2. Discharge Correlation, Auglaize River Near Defiance vs. Auglaize River Near Fort Jennings

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.

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 covered by a historical account is generally high, more often it corresponds to "flood" stage, that 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 since there is generally only one flood of such magnitude in any one year, it may also be viewed as a partial list of annual floods. In either case, the treatment is the same. Where, as is often the case, there is only one historical flood, the "maximum known", the base is the same as the flood.

All floods, both historical and those for period of record, above the high base, should be grouped together and assigned recurrence intervals in the same manner as previously discussed but based on the period of time for which they are known to be the highest.

With a large gap in magnitude between a historical flood and the highest in the period of record and no intimation of what may have happened in between, there does not seem to by any good way of combining the historical flood with the remainder of the data. In such a case the historical flood should be simply entered at a recurrence interval equal to one plus the period for which it is known to be the greatest of record.

Fitting Frequency Graphs - Having plotted a frequency diagram there appears a need for fitting a curve to the data. The fact that most stream-flow records are less than 25 years in length does not satisfy the demand for estimates of long-term destructive floods. The use of the frequency graphs for purposes of extrapolation may be dangerous, as the linear distance from 25 to 200 years seems very short on most graphs and the error of a curve fitted by whatever method may be extremely great at its outer extension. Most frequency functions, however elaborate, merely represent flexible curves with the general characteristics inherent in random observations. The data, not the functional theory, are used to define the graph and extrapolation can only be justified when the phenomena have been proven to conform to underlying law. Since no known fitted curve can serve any use in extrapolation its main purpose would therefore seem to be merely to provide a smoothing or interpolation formula. The value of an analytically fitted function therefore seems doubtful indeed. Graphical treatment only is recommended.

An example of the details of listing and plotting flood peaks is shown in a paper in *Proceedings* of the 26th Annual Meeting of the Highway Research Board (see "Use of Streamflow records in Design of Bridge Waterways" beginning on page 163), and a similar procedure will not be repeated here.

Unless a very long record is being analyzed as at major stations on large rivers, it is not advisable, except in an emergency, to prepare a frequency curve derived from one station alone. 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 otherwise of 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. In drawing the curve it should be remembered that 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. The design of the plotting paper tends to straighten out the shape of the frequency curves, but, with the above exceptions, the base data remains the best indication of the shape.

But even in such cases, if it becomes necessary to extrapolate any appreciable distance beyond the extent of the usable data, then some method other than mere extension of a curve becomes necessary. These methods might include study of ratios of the 25-or 50-yr. 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.

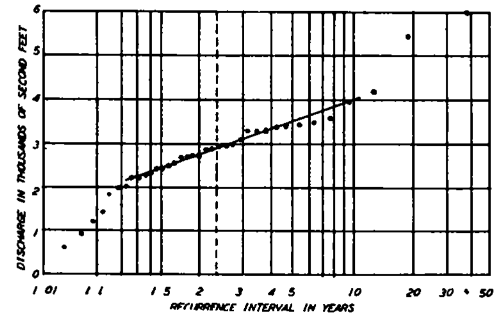

Figure 3. Preliminary Frequency Curve for Annual Floods on St. Joseph River Near Blakeslee, Ohio (1)

REGIONAL DISTRIBUTION

Flood-frequency compilations prepared in accordance with the procedure discussed above result in tables and diagrams of flood magnitude and average recurrence intervals for each station studied. These charts are not applicable directly to ungaged areas. In most cases the flood records are short and the sampling errors correspondingly large, and the records are for different periods of time. Investigations have been made of the possibilities of combining the flood data for a drainage basin or larger region, and also of relating the flood-frequency function to measurable characteristics of the drainage basin, thereby in the first instance reducing the large sampling errors, and in the second instance giving the data regional significance and so making the flood-frequency studies applicable to ungaged areas. 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 objective of such analysis is to obtain all of the relevant information from the available flood data, strengthening as much as possible the flood-frequency functions, and making available adequate methods of estimating flood frequencies for ungaged areas. In general, we are concerned only with floods not exceeding 50-yr. magnitude, and no attempt will be made to extrapolate beyond this value. In studies of this kind, it is impossible to avoid use of statistical theory. In order to aid those not thoroughly familiar with statistics, a partial list of elementary textbooks is included in the list of references.

Although flood-frequency compilations often include floods other than the maximum annual, it has been demonstrated that the partial duration series and the annual flood series give essentially identical results for intervals greater than about ten years. For this reason, annual floods only are used in these studies. When it is desirable to know how often, on the average, a stream will get above a certain discharge, as when designing a low fill across a valley on a secondary road so that the higher floods will overflow the road, the frequency curve based on the partial duration series must be used. The most simple way to do this is to convert the curve based on the annual flood series by use of the relation expressed in the table presented above; results will be entirely adequate.

Combining Records - It has been suggested that 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, five records of 20 years each could be combined to obtain a 100-yr. record, thereby increasing the accuracy of flood predictions by reducing sampling errors. Such a method requires (1) that flood-frequency characteristics be comparable and (2) 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 sq. mi. Five stations in an area essentially hydrologically homogeneous measure the same flood five times rather than measuring five floods each year; in other words, a 20-yr. period cannot yield more than a 20-yr. record, and cannot be expanded so as to become a 100-yr. period. Despite this unfortunate situation there is considerable to be gained by combining flood records in other ways.

There are two aspects to the analysis of flood frequency in a region: (1) time sampling and (2) geographic sampling. For the first, we require a sample of a large number of years, not necessarily at one station. For the second, we need a large number of stations to sample wide diversity of terrains. The two aspects need not be concurrent. For example, in a flood program one might operate a limited number of long-term base stations, supplemented by a large number of short-term or temporary stations on a large number of different kinds of streams.

Tabulation of Annual Flood Data - The first step in the flood-frequency analysis is to marshal all available annual flood data in a block table. Each station record should be identified by a number, preferably referring to a map, and the full station name and location should be given, as well as the drainage area in square miles. Four lines per record are provided for the data and computations. The calendar year data and discharge of the annual flood are listed for each water year. The date should be included in view of the possibility that the study may reveal sufficient independence between floods that some form of application of the station-year method may be utilized.

Computation of Comparable Means - Experience has demonstrated that the arithmetic average of the annual floods for one station may not be compared with the average for another station with a different length of record. The mean annual floods for each station must be comparable, and therefore must be adjusted to the same period of record. A base period may be chosen equal to any period desired for which records are available, but in general it is best

TABLE 1

DATA FOR HOMOGENEITY TEST FOR GAGING STATIONS IN MAUMEE RIVER BASIN

No	Gaging Station	Drainage Area	Mean Annual Flood, $Q_{2.33}$	10-yr. Flood, Q_{10}	Ratio $\frac{Q_{10}}{Q_{2.33}}$	$Q_{2.33}$ x 1.38	Recurrence Interval for Q of Column 7	Period of Record
1	2	3	4	5	6	7	8	9
		sq mi	sec ft	sec ft		sec ft	yr	yr
1	St. Joseph River near Blakeslee, Ohio	369	2,900	4,000	1.38	4,000	10	6
2	St Joseph River near Fort Wayne, Ind.	1,060	8,300	10,600	1.28	11,500	21	8
3	Maumee River at Antwerp, Ohio	2,049	15,200	22,300	1.47	21,000	7	37
4	Maumee River near Defiance, Ohio	5,530	48,300	68,000	1.41[a]	66,700	9	22
5	Maumee River at Waterville, Ohio	6,314	52,500	70,800	1.35[a]	72,400	12	25
6	Cedar Creek at Auburn, Ind	93	908	1,090	1.20	1,250	45	6
7	St Marys River near Willshire, Ohio	355	3,720	5,050	1.36	5,130	11	7
8	St Marys River near Fort Wayne, Ind	753	7,700	10,600	1.38	10,600	10	18
9	Bean Creek at Powers, Ohio	238	2,900	3,550	1.22	4,000	30	8
10	Tiffin River at Stryker, Ohio	444	4,080	5,480	1.34	5,630	13	17
11	Tiffin River near Brunersburg, Ohio	766	6,500	8,550	1.32	8,970	14	7
12	Auglaize River near Fort Jennings, Ohio	333	5,850	8,200	1.40	8,070	9	23
13	Auglaize River near Defiance, Ohio	2,329	27,000	42,700	1.58	37,300	6	34
14	Ottawa River at Allentown, Ohio	168	2,930	4,200	1.43	4,040	8	19
15	Ottawa River at Kalida, Ohio	315	4,800	7,300	1.52	6,620	7	5
16	Blanchard River near Findlay, Ohio	343	5,600	7,850	1.40	7,730	9	22
17	Blanchard River at Glandorf, Ohio	643	8,600	12,100	1.41	11,900	9	16
	Average Ratio - - - - - - - - - - - - - - - - - -				1.38.			

[a] The mean of these two was used to compute the average.

to use the period of the longest record.

The mean used for each frequency distribution is the graphical mean determined by the intersection of the visually best fitting frequency line with the mean line (the line corresponding to the 2. 33 years recurrence interval). The graphical mean is more stable and dependable than an arithmetic mean. This method of determining the mean in effect gives greater weight to the medium floods than to the extreme floods with large sampling errors, and for this reason is not influenced adversely by the chance inclusion or exclusion of a major flood, as is the arithmetic mean.

The graphical mean annual floods may be determined by plotting records for each station on a frequency chart and drawing the best fitted curve for a short interval on either side of the 2. 33-yr. line. However, to do this, the recurrence intervals computed for each flood of a short record must be the same as they would be if the record were continuous for the entire base period selected.

There are several ways in which recurrence intervals for a short record may be adjusted to the longer period. The most promising of these methods at present consists of computing a figure for each year of the base record for which no record was obtained. The record then would be considered complete for the purpose of determining the graphical mean (and for making a test for homogeneity). The computed figures are not true discharges and should never be so considered; they are merely computation figures.

The computed figures may be obtained by comparison of records for the short-term station with records from a long-term station, usually by means of a correlation curve based on flood peaks, plotting the peak discharge for a flood peak at one station against the peak discharge for the corresponding flood at the other station.

When the record for each station is

TABLE 2

TABULATION OF FLOOD DATA, MAUMEE RIVER BASIN

Sta No	Station and Location	Drainage Area	Water Year	1912	1913	1914	- -	1947	1948	Adjusted Mean
		sq mi								sec ft
1	St Joseph River near Blakeslee, Ohio	369	Date							
			cfs	3,400[a]	6,000[a]	2,700[a]	- -	2,750[a]	2,450[a]	2,900
			order	9	1	21		20	26	
			ratio							
2	St Joseph River near Fort Wayne, Ind	1,060	Date		Mar -			Apr 24	Feb 29	
			cfs	9,450[a]	16,500	7,900[a]	- -	6,970	6,800	8,300
			order	9	1	19	- -	27	28	
			ratio		1 99		- -	840	819	
3	Maumee River at Antwerp, Ohio	2,049	Date	Apr 2	Mar 27	May 13	- -	June 4	Feb 29	
			cfs	19,200	40,000	14,300	- -	11,600	11,900	15,200
			order	8	1	19		27	25	
			ratio	1 26	2 63	941	- -	763	783	
- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
17	Blanchard River at Glandorf Ohio	643	Date				- -	June 9	Mar 23	
			cfs	9,900[a]	28,000[a]	7,600[a]	- -	11,300	9,140	8,600
			order	12	1	21	- -	4	14	
			ratio				- -	1 31	1 06	

[a]Computed

complete, i. e., there is a figure, either of an actual measured discharge or a computed value, for each year of the total period, order numbers should be assigned and the corresponding recurrence interval computed for each year of record. The complete data should be plotted on the frequency chart and the short-interval curve drawn so as to obtain the 2.33-yr., or mean, discharge.

Computation of Flood Ratios - The adjusted mean floods, computed as described in the preceding paragraph, should be entered on the block table (see Table 2). Each individual flood that was actually measured (not the computed values) should be divided by the adjusted mean. The resulting flood ratio or flood magnitude in relation to the mean is then entered in the table. This expresses all floods in dimensionless terms and places them on a comparable basis, that is, all are measured in relation to the station mean flood for the standard period.

Test for Homogeneity of Records - The flood-frequency graphs prepared in flood-frequency compilations usually do not have the same slope at each station, but for several stations within a region the differences in slopes may not be great. If we can determine that the differences in slopes are no greater than might be expected from random errors, or vagaries of sampling, then we can combine several records to obtain an average flood-frequency curve or line that will be more accurate than any one of the individual lines, and that can be applied throughout a region.

The test set-up requires a study of the 10-yr. floods as estimated at each station. The curve, or line, drawn on the flood charts that have been prepared to compute comparable means should be extended to the 10-yr. recurrence interval. Each 10-yr. flood should be divided by the mean flood to get the 10-yr. ratio. An average of these ratios should be obtained. Then list 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.

Tentatively assume that each station represents a different sample from a single homogeneous record. If this is so, then the recurrence intervals will not differ among themselves by an amount greater than can be attributed to chance. A figure (see Fig. 4 of example) has been set up to test this

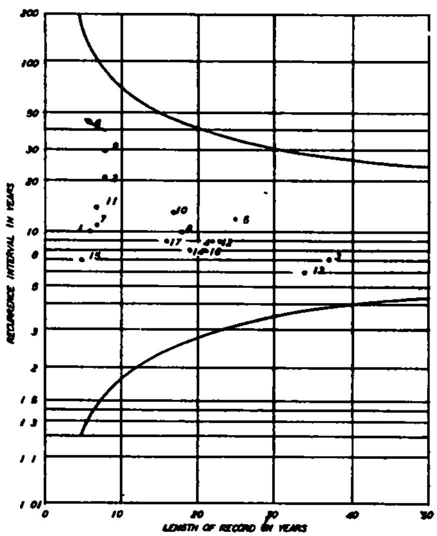

Figure 4. Homogeneity Test

supposition. It shows within what range of recurrence intervals we can expect an estimate of a 10-yr. flood to be. The range, of course, is rather great with short records, and as one would expect, the range narrows down with long records; the upper and lower limits ultimately converging on the 10-yr. interval. If all points as listed above plot within the limits, we may reasonably conclude that the records are acceptably homogeneous.

Computation of Median Flood Ratios for a Group of Stations - Where the previous tests indicate the flood characteristics at one or two stations to be quite different from the others of the group, these stations may be omitted from the others of the group, these stations may be omitted from the averaging process to be described. In general, however, don't be in a hurry to discard a record; the apparent lack of homogeneity may be due to poor record or similar reasons.

There are also cases where stations may be too homogeneous, as for example, stations close together on the same stream. In general, if two stations on the same stream do not differ by more than 25 percent in flow, the one with the shortest record may be dropped from the list. Where the records do not overlap or are only partly overlapping, the records may be combined to make one long record for the flood-frequency studies.

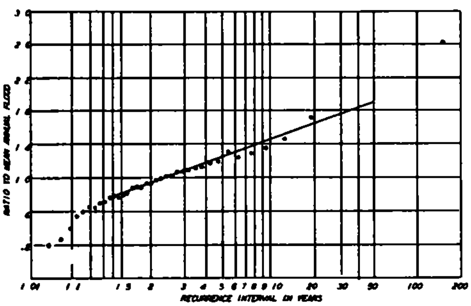

Figure 5. Composite Frequency Curve for Annual Floods in Maumee River Basin

All stations satisfying the above requirements should be grouped for the purpose of computing average recurrence intervals, applicable for the region. For every station, list each flood ratio in order of magnitude for those floods actually measured (not for the computed values); usually there will be many blank spaces in the table.

The table will show one or more flood ratios for the measured floods having an order number of one, one or more flood ratios for the measured floods having an order number of two, and so on for each order number up to the highest. For each order number the median flood ratio should be recorded; this median is the mid value of an odd number of events or the mean of the two central values of an even number of events. The recurrence interval corresponding to each order number should also be listed (see Table 3 of example).

Each median flood ratio should be plotted to its corresponding recurrence interval on a frequency chart and an average frequency curve drawn (see Fig. 5 of example). This curve, showing flood discharge in ratio to the mean annual flood, is based on all significant discharge records available and may be considered as representing the most likely flood frequency values for all places in the region.

Extrapolation of Flood-Frequency Graphs - The primary purpose of flood-frequency studies is to provide information on the magnitude of a flood at a specified location and with a specified frequency or recurrence interval. This specified recurrence interval normally is beyond the length of the record. Involved, extensive, and apparently never-ending theoretical studies of flood frequencies have generally attempted to produce a plotting system that gives a straight line plot of frequencies so that extrapolation can be made with diminished error. The sampling errors of flood records are such that all of these attempts are more or less in vain. For example, the graph paper and plotting system herein recommended has much in its favor, both theoretically and practically, as far as the abscissa or recurrence interval scale is concerned. The use of a linear ordinate or discharge scale, however, is entirely arbitrary, having no theoretical support, and has nothing to recommend it except that many flood records approach straight lines by the use of a linear discharge scale; it may be that the ordinate scale should be logarithmic or some other function of the discharge. For these reasons long extrapolation on a flood-frequency graph is extremely dangerous.

Reasonable requests for data on flood frequencies must be met if practical value is to be given to the analysis. For many purposes, as for design of highway structures, the need is for knowledge of floods having a frequency of only about 50 years. An extrapolation to this interval usually is not great

TABLE 3

PLOTTING POSITIONS FOR MEDIAN FLOOD RATIOS, 1912-48, MAUMEE RIVER BASIN

Order Number	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	Median Ratio	Plotting Position
Col 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		1 99	2 63							1 71			3 33			2 86		2 63	38 0
2	1 88		1 72		1 58			1 74		1 44	1 54	1 50	1 78	1 76		1 66	1 73	1 72	19 0
3		1 28	1 48		1 54				1 26	1 39		1 45	1 67	1 72		1 46	1 45	1 46	12 7
4	1 37		1 39		1 36				1 23			1 35	1 46			1 38	1 31	1 36	9 50
5			1 39			1 18				1 26			1 44			1 33	1 27	1 30	7 60
6		1 22	1 36		1 24					1 24			1 44	1 32		1 20		1 24	6 33
7			1 36					1 22				1 26	1.40					1 31	5 43
8	1 18	1 20	1 26			1 11							1 37					1 20	4 75
9			1 23	1 18					1 13			1 21				1 14		1 18	4 22
10			1 21		1 14	1 08	1 15	1 06			1 13		1 27					1 14	3 80
11			1 13		1 10	1 08						1 19	1 19			1 10		1 12	3 45
12			1 10		1 10		1 06					1 17	1 16	1 07				1 10	3 17
13			1 08					1 05		1 07			1 16					1 08	2 92
14			1 03	1 08								1 08	1 15	1 06		1 07	1 06	1 07	2 71
15			1 03		1 04	1 01					1 04		1 02	996				1 02	2 53
16			1 02		1 02	1 01				1 00		1 06	1 01			1 03		1 02	2 38
17			993		1 02			974					989			1 00	996	994	2 24
18			993		998			961		980				955			977	978	2 11
19			941		962							923	944			964		944	2 00
20			941		949				996	973			930	915		893		941	1 90
21			927					929				885	874	894		875		890	1 81
22			914		934							885	833				872	885	1 73
23			882	895				860		881			826	884				882	1 65
24			829		880		788	852				783	815	864		826	845	829	1 58
25			783		872					804			792	853			779	798	1 52
26			763		866		736			789		750	778		798	723		770	1 46
27		840	763		848					784	861	735	775	805				794	1 41
28		819	744							767								767	1 36
29			744		789			669				670	760				687	716	1 31
30	769	801	710				629	659				656	696	758				703	1 27
31			698		720		619	644	607				667	751		589		656	1 23
32			671		697			627				615	644	744	715		523	658	1 19
33	648	731	581	622				576		490	680	601	637			511	519	601	1 15
34			574		596			476	552			547	559	693	464	502	453	550	1 12
35			494		488			455	479	355	357	395	448	423	290	275	395	409	1 09
36	317		256		240			386	297	289		256	307	314	265	214	221	277	1 06
37			222		206		152	260			154	236	194			171		200	1 03

and may be made with considerable confidence.

Estimation of Flood Frequencies on Ungaged Areas - It may be assumed that a flood-frequency distribution computed as above may be applied to all streams in the region with drainage areas within the range covered by the stations used in deriving the distribution. In order to apply the flood-frequency distribution to ungaged areas it is necessary to estimate the mean flood for each area. This involves a correlation analysis of the observed mean floods with drainage basin characteristics.

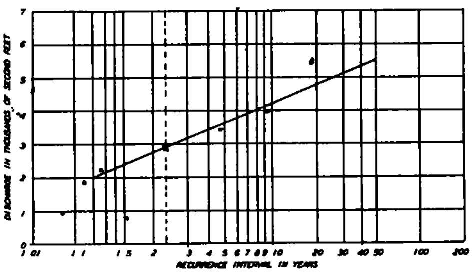

Figure 6. Frequency Curve for Annual Floods on St. Joseph River Near Blakeslee, Ohio (1)

Assuming that the region is hydrologically homogeneous the factors which affect the mean flood are area, topography, shape of the drainage basin, channel storage, and doubtless others. Of these the most important is area, the factor most readily available. Measuring the other factors is more difficult, and unless good topographic maps are available may be impossible. Channel storage undoubtedly has an important effect, but cannot be directly measured, and must be considered as indirectly accounted for by other factors, such as slope and shape.

Many methods of correlation analysis are available; Ezekiel[2] gives a full discussion of this problem. Kinnison and Colby[2] describe a method of correlating topographic characteristics with discharges which they applied successfully to New England flood data. In selecting data for this sort of analysis, emphasis should be given to diversity of terrain and to small streams. Length of record is of minor consideration, provided that the mean flood is adjusted to the base period. The main point to accent is to include a large number of basins, so as to sample as wide differences in physical characteristics as possible. The basins included may cover a much wider region than were grouped in the frequency analysis.

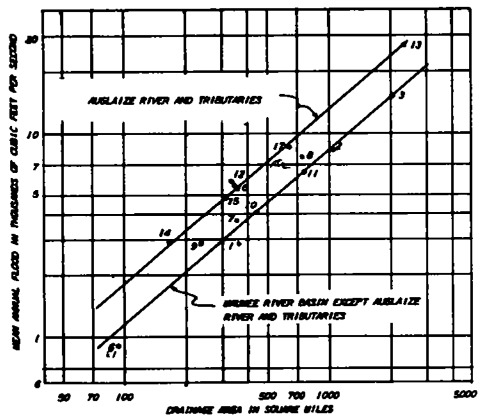

Figure 7. Mean Annual Flood, Maumee River Basin

Whatever method of correlation is used the resulting formula will not fit the data perfectly. The errors may be related to some factor not included in the correlation, such as permeability of the soils, differences in rainfall, or channel storage, as well as random errors. In some regions, area alone may furnish the best correlation that can be made; in other regions, other factors might well be included, as area of lakes and swamps.

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. Plot, usually to log-log scales, mean annual flood discharge as ordinate and drainage area as abscissa. Fit a mean curve graphically to the data (see Fig. 7 of

[2]See list of references at the end of this paper.

example). From this curve can be read the mean annual flood discharge for any stream in the region.

The mean annual floods must be multiplied by the flood ratios previously computed to obtain the discharge corresponding to a selected frequency. If desired, a complete frequency curve can be produced by plotting discharges for several frequencies and drawing the curve they define.

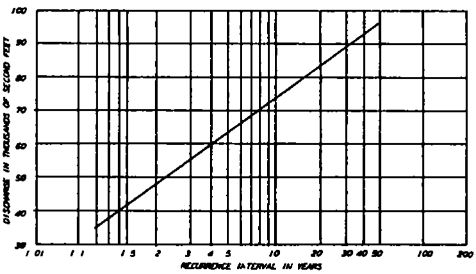

Figure 8. Frequency Curve, Main Stem Maumee River, Below Defiance, Ohio - Note: This curve is based on a drainage area of 6,000 sq. mi.; for other size of areas from 5,500 to 6,600 subtract or add 1 percent for each 100 sq. mi.

Summary of Procedure - The step by step procedure for determining regional flood frequencies is as follows:

1. Tabulate flood data for all gaging stations in the region having a record of about five years or more. List maximum annual floods.
2. Review station history and make careful study of the stage-discharge relationship.
3. Select base period for study. Usually this will be the period of the longest record.
4. Compute discharge figures for missing years for each station.
5. Number floods for each station in order of magnitude, numbering the greatest flood 1.
6. Compute recurrence intervals by formula: R. I. = (N+1)/M.
7. Plot discharge vs. recurrence intervals and draw frequency curve to get the 2.33 and 10-yr. floods.
8. Test for homogeneity.
9. List annual floods in block table for homogeneous stations, and record the mean annual, or 2.33-yr. flood.
10. Compute ratio of annual flood to mean flood for each year of actual record.
11. Tabulate flood ratios, listing ratios for each order number on one line or in one column.
12. Determine median flood ratio for each order number, and record corresponding recurrence intervals.
13. Plot median flood ratios vs. recurrence intervals and draw composite frequency curve. In general, do not extrapolate above 50-yr. frequency.
14. Plot mean annual floods vs. drainage areas. Draw curve, or curves, to show relation applicable for region.
15. Determine flood frequency for any place in region from curves of items 14 and 15.

EXAMPLE

As an example of the application of the method described in this paper, a regional flood frequency analysis is presented for the Maumee River Basin, which drains parts of Ohio, Indiana, and Michigan.

A map of the Maumee River Basin is shown as Figure 1. The numbers on map show location of gaging stations used in the analysis and they may be identified by reference to Table 1.

The procedure outlined below is referenced to the items numbered in the "Summary of Procedure" section.

1. Annual flood peaks, in cubic feet per second, were listed for the 17 gaging stations in the basin having a record of five years or more. The maximum flood peak was listed for each water year, or year ending September 30.

2. A review of the stage-discharge relation was made and discharges were recomputed for several stations where more recent high-water discharge measurements indicated the originally used rating curve, which was usually extrapolated, was in error. Slight changes were made in order to make each record consistent within itself.

3. The base period selected was

1912-48. This is the period of the longest record, that for Maumee River at Antwerp, Ohio.

4. Discharge figures for years of missing record were computed by correlating the discharge at one gaging station with that for another, beginning with the complete record for the Antwerp station. Figure 2 shows the correlation curve for Auglaize River near Defiance vs. Auglaize River near Fort Jennings.

5. Order numbers were assigned from 1 to 37, numbering the largest flood 1.

6. Recurrence intervals were computed as (N+1)/M, where N is number of years of record, in this case 37, and M is the order number.

7. Discharges were plotted to their proper recurrence intervals on frequency charts and the frequency curve drawn from below 2.33 years to the 10-yr. interval. No attempt was made to put the curve through the low or the high points, as only the middle frequencies were needed. An example of this plot is shown for St. Joseph River near Blakeslee, Ohio, as Figure 3.

8. Data for homogeneity test are listed in Table 1. Data for columns 4, 5, and 8 were taken from frequency charts as discussed under item 7; data in other columns were taken from original records or computed.

The test for homogeneity was made by plotting data in columns 8 and 9 of Table 1, as shown in Figure 4. All data plot within the limits, therefore these records are considered to be homogeneous and were combined in one group.

9. Table 2 shows in outline the block tabulation of flood data from each gaging station in the basin. The mean discharge adjusted to the base period 1912-48, listed in the last column, were taken from the individual plots as discussed under item 7, and are listed in column 4 or Table 1.

10. The ratio of annual flood to mean flood for each year of actual record is shown on the fourth line of data for each station in Table 2.

11. Table 3 shows flood ratios for all floods of record listed in Table 2, with floods having the same order number listed on one line.

12. The median flood ratio for each order number and the corresponding recurrence intervals are listed in columns 19 and 20 of Table 3.

13. Data from columns 19 and 20 of Table 3 were plotted on a frequency chart and a curve fitted graphically. This composite frequency curve shows the ratio to the mean annual flood for recurrence intervals up to 50 years and is presented as Figure 5.

In plotting data from Table 3, the highest ratio was plotted at 150 years rather than at 38 years, as is given by the data. An examination of the records shown in Table 2 shows that at every station the March 1913 flood is the highest for the period. Records for the Miami River at Dayton, Ohio, a stream that heads just across the watershed to the south from Maumee River, shows that the March 1913 flood was the greatest at least since 1805 (see: "Floods of Ohio and Mississippi Rivers," January and February 1937; Geological Survey Water Supply Paper 838, page 654); local knowledge which began about 200 years ago does not include a greater flood. The rainfall causing the March 1913 flood in this area centered over the Maumee-Miami watershed (see: "The Ohio Valley Flood of March-April 1913"; Geological Survey Water-Supply Paper 334, plate III) and it has been assumed that this flood in the Maumee Basin would be of about the same frequency as in the Miami Basin; this assumption is also supported by local knowledge. The period 1805-1948 is 144 years but this has been rounded to 150 as it is not an exact figure. The curve was not drawn to pass through this upper point as its correct position is still unknown; the 150-yr. plotting certainly is more nearly correct than the 38 years indicated by the period of continuous record.

The curve as drawn may not actually fit every location on every stream in the basin but it does reflect average conditions and is a good representation of probable values. A check on how well this average curve fits the data for the individual gaging stations used in the analysis can be made by plotting the curve with a plot of the measured discharges at each station; this has been done for the 17 stations studied and the agreement is good, especially for those stations having the longer record. As an example, the curve for St. Joseph River near Blakeslee, Ohio, is shown as Figure 6.

14. The mean annual floods were plotted against drainage areas (columns 3 and 4 of Table 1), except for the two lower stations on the main Maumee River, and are shown on Figure 7. These points define two curves, and a study of the records shows one curve, the highest, is applicable for the Auglaize River sub-basin and the other curve is applicable for the rest of the region.

It is believed the reason the curve for the Auglaize River Basin is different and greater than for the rest of the region is because the normal direction of travel of storms, from southwest to northeast, is down stream for nearly all of the tributaries. This would tend to produce greater floods in this stream, and is a condition that does not apply to the other sub-basins in the Maumee River Basin.

The curves shown on Figure 7 do not extend below 75 sq. mi. and should not be extrapolated to smaller areas. The smallest area gaged is 93 sq. mi. and to extrapolate further would be unwise. When data becomes available from smaller areas the curves can be extended downward and their value increased.

The lower Maumee River was not included with the other streams as the drainage area is much larger, and a different treatment will be shown below.

15. To determine the magnitude or frequency of recurrence of a flood in the Maumee River Basin, proceed as follows: (a) locate the stream where it is desired to determine the magnitude of, say, the 25-yr. flood; (b) measure the drainage area, in square miles, from topographic or other map; (c) from Figure 7 read the mean annual flood in cubic feet per second, corresponding to the drainage area, being careful to use the proper curve; (d) from Figure 5 determine the ratio of the 25-yr. flood to the mean annual flood; and (e) multiply the two figures obtained to give the 25-yr. flood at the point in question.

As an example, assume it is desired to know the 25-yr. flood from a 500 sq. mi. drainage area in the Auglaize River sub-basin. From Figure 7, the mean annual flood is found to be 7,200 sec.-ft.; and from Figure 5, the ratio of the 25-yr. flood is found to be 1.71; the 25-yr. flood, therefore, is 12,300 cfs.

For the main stem of the Maumee River below Defiance the drainage area varies from about 5,500 to 6,600 sq. mi. and the flood discharges vary about 1 percent for each 100 sq. mi. One frequency curve may be used for this main stem, presented as Figure 8, that will have a maximum error of about 5 percent. For more accurate use, results obtained from Figure 8 should be reduced or increased one percent per 100 sq. mi. for areas less than or more than 6,000 sq. mi.

ACKNOWLEDGMENTS

The flood frequency method presented in this paper is the result of several years of study by a number of engineers of the Water Resources Division of the United States Geological Survey. Special recognition is given to Mr. W. P. Cross, Hydraulic Engineer, Surface Water Branch, Columbus, Ohio, and Mr. W. B. Langbein, Hydraulic Engineer, Technical Coordination Branch, Washington, D. C.

REFERENCES

Statistics

1. Arkin, Herbert, and Colton, Raymond, "An Outline of Statistical Methods," 220 pp., Barnes and Noble, New York, 4th Ed., 1939.
2. Snedecor, George W., "Statistical Methods," 485 pp., The Iowa State College Press, Ames, Iowa, 4th Ed., 1946.

3. Hoel, Paul G., "Introduction to Mathematical Statistics," 258 pp., John Wiley and Sons, Inc., New York, 1947.
4. Fisher, R. A., "Statistical Methods for Research Workers," 345 pp., Oliver and Boyd, London, 10th Ed., 1946.
5. Ezekiel, Mordecai, "Methods of Correlation Analysis," 522 pp., John Wiley and Sons, Inc., New York, 1945.
6. Rider, Paul, "An Introduction to Modern Statistical Methods," John Wiley and Sons, Inc., New York, 1939.

Floods

7. Hazen, Allen, "Flood Flows," John Wiley and Sons, New York, 1930. (A general treatise on the subject of floods, with special emphasis on the determination of frequency and magnitude by the use of logarithmic skew-frequency curves.)
8. Fisher, R. A., and Tippett, L. H. C., "Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample," *Proc.* Cambr. Philo. Soc., Vol. 24, pp. 180-190, 1928. (Develops frequency distributions of maximum values, subsequently applied by Gumbel to floods.)
9. Gumbel, E. J., "Floods Estimated by Probability Method," *Engrg. News-Rec.*, June 14, 1945.
10. Powell, R. W., "A Simple Method of Estimating Flood Frequencies," *Civil Engineering*, pp. 105-6, Feb. 1943. (Introduces the Gumbel method to engineers.)
11. Foster, H. Alden, "Theoretical Frequency Curves and their Application to Engineering Problems," *Trans.* Am. Soc. C. E., Vol. 87, pp. 142-203, 1924. (The application of the family of flexible skew-distribution curves, devised by K. Pearson, to frequency distributions of annual floods. Fitting of these curves requires selection of type and calculation of mean and coefficients of variations and skew.)
12. Jarvis, C. S., "Floods in the United States, Magnitude and Frequency," U.S. Geological Survey, Water-Supply Paper 771, 1936. (A comprehensive source book of facts on flood frequency. Gives an outline of the many different methods for determining flood frequency.)
13. Slade, J. J., "The Reliability of Statistical Methods in the Determination of Flood Frequencies," U.S. Geological Survey, Water-Supply Paper 771, pp. 421-432, 1936. (An analysis of the errors inherent in calculations of the mean, coefficient of variation, and the coefficient of skew with small samples.)
14. Cross, W. P., "Floods in Ohio, Magnitude and Frequency," Ohio Water Resources Board, Bull. 7, 1946.
15. Creager, W. P., Justin, T. D., and Hinds, J., "Engineering for Dams," Vol. 1, John Wiley and Sons, New York, 1944.
16. Rantz, S. E., and Riggs, H. C., "Magnitude and Frequency of Floods in the Columbia River Basin," U.S. Geological Survey, Water-Supply Paper 1080, pp. 317-469.