U. S. DEPARTMENT OF THE INTERIOR
Geological Survey
Water Resources Branch

INSTRUCTIONS FOR FLOOD FREQUENCY COMPILATIONS

Prepared July 1946
Revised September 1946
Revised April 1947
Revised______________

HANDBOOK FOR HYDROLOGISTS

These instructions are designed to facilitate work in this field, but are not to be considered as final and complete. They will be revised as needed to incorporate new ideas and experience gained in this field. For that reason, if use is planned any great period after the date shown on the cover, write the Chief Hydraulic Engineer for advice as to changes. Comments are always welcome.

OUTLINE OF PROCEDURE

1. Prepare station description (see p. 7).

2. List flood discharges (see sample form 9-179 following p. 7). Note conditions for listing partial duration series ("peaks above a base"), see p. 5.

3. Compute mean annual flood on form 9-179 and enter in space provided at foot of column. Show dates.

4. Compute plotting positions of annual floods and partial duration series, (see pp. 4-5).

5. Plot annual floods on Gumbel chart (see sample).

6. Plot partial duration series on semi-log chart (see sample).

FLOOD-FREQUENCY COMPILATIONS

The purpose of these instructions is to establish uniform procedure and consistent practice for the Survey in making what are commonly called "flood-frequency compilations." Studies of flood frequencies have been found especially helpful in problems involving economic considerations. Design of bridge clearances, channel capacities and roadbed levels are common problems among the many that involve the determination of flood discharge. Where costs must be balanced against flood damage or liabilities arising from failure, there is a place for the study of flood frequency. The records of the Geological Survey are very helpful in such practical problems.

The flood-frequency method has encountered considerable criticism, largely, it is believed, because of abuse. The method has little place in determining maximum limits of flood design (i.e. "the maximum possible flood"). With the ordinary stream-flow record, of say 2_-year length, errors of sampling introduce large errors in judging the magnitude of floods of greater than about 15-year magnitude. However, properly computed and conservatively interpreted, flood-frequency analysis can be a valuable hydrologic tool.

The subject has been an attractive one to many students and it accordingly has benefited by their voluminous writings. The viewpoints and theories expressed, although instructive, have not always been consistent. This guide is a brief digest of what seems to be the most practical of the manifold methods available in this field. Sufficient explanation is given to enable the engineer to know what he is doing. The guide does not restrict or discourage the study and experiment with methods other than those recommended. The Washington office should be informed, however, of proposed modifications in procedure.

There are three major aspects to the problem: viz. (1) The kind of flood data to be studied (2) Plotting positions and (3) Fitting frequency functions.

In the following, brief reference is made to several diverse ideas on these points and specific recommendations are made in so far as Survey practice is concerned.

General

The first step in beginning a flood compilation is to select the gaging station. Gaging stations to be considered for this kind of study should have at least 15 years of record. Storage or other artificial factors which would tend to modify flood discharges significantly should be a minimum. In all cases include a statement as to total usable storage capacity in basin above gage.

Floods are to be listed in chronologic order. Peak stages and discharges are both to be listed. Stages are included for their own intrinsic value and to enable comparison with discharges as a rough 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.

It is to be noted that peaks only are included in this table. Daily discharges are excluded even though W.-S. P. 771 shows daily discharges almost exclusively. That situation, it is believed, 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 sensibly the same.

A sample tabulation (form 9-179) is included as an appendix to these instructions.

The second step is to review the station history, especially the stage-discharge rating. The determination of the proper rating and peak discharges requires careful study and considered judgment. This review should be performed by persons experienced in field and office stream-gaging practice. No specific instructions can be given but 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 measurement and neglect to revise past records. In any case do not use published data without examining the record for consistency. Stations with poorly defined rating extensions might well be omitted.

Where previously published figures are revised or adjusted for flood-frequency compilations, the relation of the new figures to previously published figures must be established. If the changes are so large that the regular rules for revising records in the annual surface water reports will hold, the revisions should be published in those reports. In that event, previous figures are superseded. On the other hand, the new figures often represent minor changes from previously published figures made solely for the purpose of obtaining consistency among the flood-peak discharges being studied. If it is carefully explained that these changes are made for the purposes of flood frequency compilations only, they need not be considered as revisions. In that event the new figures do <u>not</u> supersede previously published figures and the records in annual surface-water reports are <u>unaffected</u>.

Kind of data to be studied

A. Annual floods. An annual flood is defined as the highest 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.

B. Partial duration series ("floods above a base"). This objection is met by listing all floods above a selected base without regard to number within any given time period. The floods are numbered with respect to size, beginning with the highest as number 1. 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 followed in the selection of the floods is that each peak be individual, i.e., be separated by substantial recession in stage and discharge. (See section on "detailed plotting procedures")

An objection to the use of the partial flood series is that the floods listed may not be fully independent events, i.e., one flood sets the stage for another. A related objection is that when the listed floods are so closely consecutive, the flood peaks may actually be one, as the damage was caused by the highest; the associated peaks may only have indirect or secondary effects on the losses.

In a subsequent section, the differences between the two kinds of data are largely resolved. However, it is good practice to work up flood data both ways.

Meaning of recurrence interval.

In this guide, floods are classified as to recurrence interval in years. The term recurrence interval, T, is variously defined viz.

(1) Average interval in years between floods of specified magnitudes.

(2) Reciprocal of the probability that a flood of given magnitude will occur in a single year.

Definition (1) is obviously $T = (t_1 + t_2 + t_3 + \ldots\ldots)/F$ where t_1 etc. represent the recorded interval between floods of specified magnitude and F = total number of these floods. However, where a record is long relative to T, $t_1 + t_2 + t_3 + \ldots$ may be taken equal to N, where N equals the number of years of record. For short records $t_1 + t_2 + t_3 + \cdots$ differs significantly from N. For the highest flood of record, for example, there is no way in which to evaluate any of these intervals.

Since the annual probability of a flood of given magnitude is f/N, definition (2) follows from definition (1).

It must be emphasized that these definitions apply only to large samples. For statistically small samples such as encountered in the usual flood record, other considerations as described below are usually encountered in computing recurrence intervals.

Plotting positions

After collecting the data according to one or both of the above specifications, it is customary to number the discharges in order of magnitude. The problem then is to determine what recurrence intervals these discharges represent.

There are really two parts to the problem of plotting positions. Having the discharges in order of magnitudes, there is known (1) the relative distribution of floods in (2) a given period of years. One might, entirely apart from the length of the record prepare a graph showing percent of time that floods of given magnitude might recur, based on the sample. The second 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.

Consider the array to contain 20 items numbered from 1 for the highest to 20 for the lowest. The highest might be reported as equalled or exceeded 1/20 or 5 percent of the time and the lowest as equalled or exceeded 20/20 or 100 percent of the time, a perfectly plausible procedure. But now suppose the array were renumbered from 1 as the lowest to 20 as the highest, an equally plausible procedure--now, our percentages become 1/20 or 5 percent for the lowest and 100 percent for the highest. Obviously the sum of the percentages that a flood can be exceeded, or fallen short of, must equal 100.

To meet this anomalous situation, Hazen proposed a compromise. According to the first system of numbering the highest flood can be exceeded 5 percent of the time, but according to the second system this percentage must be 0 (100 - 100 = 0). Hazen proposed that the average of 2-1/2 percent be used, according to the following formula, $p = (m-\frac{1}{2})/n$ where m is the relative magnitude of the item, n is the total number of items and p is percent of time. This formula will produce consistent results throughout the range, whichever system of numbering is used.

However, Gumbel (Engr'g News-Record, June 14, 1945) points out that the maximum floods in several samples of n years each will, like any other statistic, have a characteristic distribution. He says that the proper plotting position of a maximum flood is at the modal (i.e. most common) value of that distribution. Moreover the m'th value of these several samples will also have their own characteristic distribution. The plotting position for each should correspond to the modal value of its distribution.

The matter would be readily solved if we had a number of samples of n years each; however, we only have one, and therefore are in no position to determine modal values of the m'th magnitudes. Gumbel by a special logic derives a plotting position that assumes that the single sample available corresponds to the modal, or most probable value. He presents special formulas and tables for the plotting positions of the highest and lowest value in an array. To simplify the work Gumbel specifies that the remaining plotting positions be interpolated between the highest and lowest values.

Gumbel makes a very good case for his plotting positions. The uncertainties and vagaries in sampling, however, do not seem to justify the expenditure of time for this purpose. Moreover, the method is inapplicable to the partial-duration series.

Consider the array to consist of n items. In a list of n items there are n + 1 intervals. In a perfectly-ordered array these intervals would represent equal probabilities. Thus a discharge above the highest in the array would have a chance of occurrence equal to that between say the 5th and 6th items, even though in terms of discharge the intervals may be quite unequal. There is the same chance that the discharge will be less than the lowest in the array, and so on. Therefore the probability that the top item will be exceeded is $1/(n + 1)$, the probability that the next to the top item will be exceeded is $2/(n + 1)$ and so on. The probability that the lowest item will be exceeded is $n/(n + 1)$, and the probability that the discharge will be greater than zero is $(n+1)/(n + 1)$ or unity.

The formula $p = m/(n + 1)$ moreover gives results acceptably in conformance with those computed by Gumbel's theory; it is simple to compute and it is applicable both to annual flood data and the partial-duration series. Accordingly, for the purpose of this handbook, positions for plotting will be computed on the basis of the formula $p = m/(n + 1)$. However, since interest ultimately centers in recurrence intervals in years, the plotting chart is prepared on that basis.

Recurrence intervals 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 one. In the case of annual floods $n = N$ but in the partial duration series $n > N$.

Detailed plotting procedures

Annual flood peaks--List the highest observed peak in each water year in chronological order. Show calendar-year dates. Only complete years of stream-flow records can be included but historical flood data can also be included to the extent indicated below. The peaks (excluding historical data) should be numbered in order in magnitude beginning with 1 for the highest. Compute recurrence interval in years by the formula $(N + 1)/M$ where N equals the number of years in the record and M equals the order of relative magnitude as assigned. Plot on Gumbel chart paper. (See example.) The greatest known flood will plot at a recurrence interval equal to one plus the number of years in the period in which it occurred.

1. Either use longest continuous period only, or use all complete years of stream gaging. No selection may be made of portion of records to be used except to the extent of using the longest continuous period.

2. In the annual floods a record may begin say in April just a few days prior to 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 flood. The recorded flood may then be accepted as an annual flood.

3. Fragmentary historical flood data are selective and may not be used, except the highest known as described below.

It may be that from historical evidence, the highest flood in the period of record, is also known to be greatest for many years preceding the period of record. In that event, it should be plotted at a recurrence interval equal to one plus the period for which it is known to be the highest flood.

Another situation may exist where a great flood prior to the period of record, whose discharge is known, is subsequently exceeded by a flood within a period of record. Recurrence intervals are computed as follows:

Hypothetical example: Assume a discharge of 1,000 cfs. in 1850; the record begins in 1910, but the above stands as "maximum known" until 1938 when a discharge of 2,000 cfs. was recorded. Hence plotting positions (up to and including the 1945 flood) would be

max. flood in 95-year period	$\frac{95+1}{1}$	96 years	2,000 cfs.
2nd highest in 95-year period	$\frac{95+1}{2}$	48 years	1,000 cfs
2nd highest in 35-year period	$\frac{35+1}{2}$	18 years	800 cfs
3rd highest in 35-year period	$\frac{35+1}{3}$	12 years	600 cfs

etc.

Annual flood data should be plotted on Gumbel graphs. The discharges are plotted to a linear scale as ordinate. The abscissa (scale of recurrence-intervals) is specially graduated according to Gumbel's "theory of largest values."

Gumbel charts: For the general purposes of flood-frequency graphs the kind of graduations on the paper is of no great importance. The graph as plotted is to be used only for purposes of interpolation; extrapolation is not to be considered under any circumstances in any published report.

However, it is desirable to have uniformity, and if a choice is to be made the Gumbel chart has much to offer. Flood discharges plotted against recurrence intervals on this paper approximate a straight-line graph.

The forms (form 9-179a) can be furnished on letter request to the Chief Hydraulic Engineer.

Partial duration series--List all peaks regardless of date of occurrence above a selected base. Ordinarily this base should be chosen so that the number of peaks is at least equal to the years of record, but not more than three or four floods per year.

A peak shall be defined as a discharge which significantly exceeds the preceding and following discharges. It should be at least 25 percent greater than the adjacent troughs, and in general each peak should be associated with separate and distinct meteorologic events. Under most conditions this would mean that peaks should be separated by a period of at least a day.

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 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 3-cycle semi-log graph paper (K & E 358-71) (see example); using the linear scale (ordinate) for the discharge data, and the logarithmic scale (abscissa) for recurrence intervals. Connect the plotted points by straight lines.

Fitting frequency functions

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, 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. 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. Extrapolation can only be justified when the phenomena have been proven to conform to underlying law. Philosophically speaking, one set of data cannot both be used for deriving and proving a basic theory. 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 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 contemplated for Survey compilation reports. The Gumbel chart is recommended for annual floods, because it is based on a logical a priori theory of flood occurrences. Flood data should approximate a straight line on this paper. Semilog graphs should be used for partial duration series, with discharge on the linear scale and recurrence intervals on the logarithmic scale. The several plotted points should be merely connected by straight lines.

Comparisons between annual floods and the partial duration series (floods above a base)

Plottings of flood data by either the annual flood method or the partial duration series will show equivalent results for the higher or less frequent floods. For the lower floods the annual flood graph will be consistently below that of the partial-duration series. This relation is clearly shown in the case of the one-year flood. It is the lowest of the annual floods.

The one-year flood by the usual plotting procedure in the partial-duration series becomes the flood where M = N. Obviously the first is less than the second. There is a systematic relationship between the two, however, that can be derived from basic statistical theory.

Recurrence intervals in years

Annual floods	Partial duration series
1.1	0.41
1.25	.62
1.5	0.91
1.75	1.18
2.00	1.45
2.50	2.0
5.00	4.6
10.00	9.5
15.00	14.5
20.00	20.00
100.00	100.00

An annual flood is the maximum of all floods in a given year whereas a flood in the partial duration series is selected as exceeding a certain base and without reference to the number of other floods in the year. However, since a large flood is apt to outrank any other flood in the year in which it occurs, the recurrence intervals of great floods are closely the same on both scales.

The above table was computed on the assumption of floods occurring as completely independent events. However, the effect of interdependence is such as to make the items in the second column somewhat high.

There is an important distinction in meaning as 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, 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 two become sensibly numerically the same.

The annual flood series might, for example, be used in design of a bridge which is apt to be destroyed only once in a year. In this case the flood to be considered is the highest flood in a year. Other floods, although they may exceed ranking floods in other years, will be safely passed. However, consider a highway which will be flooded but not necessarily destroyed by any flood, or if damaged, will be rapidly repaired and thus soon again exposed to risk. In this case we should employ the partial duration series.

The two methods give essentially identical results for intervals greater than about ten years. Since most designs are for intervals greater than this, it is apparent that, from a practical standpoint, either method is satisfactory, although perhaps the simplicity of the annual flood method makes it attractive.

BIBLIOGRAPHY

Hazen, Allen, Flood flows, 195 pp., John Wiley, 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.

Gumbel, E. J., Floods estimated by probability method, Engr'g. News-Rec., June 14, 1945.

An elementary outline of the authors' papers on flood-frequency. Defining a flood as the largest daily discharge in a 365-day period and considering the daily discharges as an unlimited statistical variable of experimental type, Gumbel derives a form of flood-frequency distribution, that has only two parameters, viz. mean, and coefficient of variation, Gumbel also describes a method of determining plotting positions of observed flood discharges. Gumbel's methods apply only to annual floods.

Foster, H. Alden, Theoretical frequency curves and their application to engineering problems, Trans. A. S. C. E., pp. 142-203, vol. 87, 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.

Jarvis, C. S., Floods in the United States, magnitude and frequency, 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.

Slade, J. J., The reliability of statistical methods in the determination of flood frequencies, Water-Supply Paper 771, pp. 421-432, 1936. An analyses of the errors, inherent in calculations of the mean, coefficient of variation, and the coefficient of skew with small samples.

Cross, W. P., Floods in Ohio, magnitude and frequency, Ohio Water Resources Board, Bull. 7, 1946.

Creager, W. P., Justin, T. D., and Hinds, J., Engineering for dams, vol. 1, John Wiley and Sons, New York, 1944.

Powell, R. W., A simple method of estimating flood frequencies, Civil Engineering, pp. 105-6 Feb. 1943.

W.B.L.

MUSKINGUM RIVER BASIN

Licking River at Toboso, Ohio

Location.- Lat. 40°03'26", long. 82°13'12", highway bridge at Toboso, Licking County, 3 miles downstream from Rock Fork.

Drainage area.- 672 square miles.

Records available.- September 20, 1921 to September 30, 1945.

Gage.- Non-recording gage read once daily prior to September 20, 1929, recording gage thereafter. Datum of gage is 744.84 feet above mean sea level, adjustment of 1912.

Stage-discharge relation.- Defined by current-meter measurements below 24,000 cfs.

Maximum flood of record.- 28,900 cfs, January 25, 1937.

Historical data.- Flood of March, 1913, reached a stage of 20.0 feet, present datum (discharge 35,000 cfs, computed by Muskingum Watershed Conservancy District).

Remarks.- Gage-heights for period of non-recording gage are from graphs based on gage readings. Flow slightly regulated by Buckeye Lake on South Fork, usable capacity 27,300 acre-feet.

19336

9 - 179
July 1946
Experimental
Flood Data
2088

UNITED STATES DEPARTMENT OF THE INTERIOR
GEOLOGICAL SURVEY
WATER RESOURCES BRANCH

File

Flood data for Licking River at Toboso, Ohio

Drainage area 672 sq. miles. Period of record 1921-1945

Gage Non-recording to Sept. 1929, recording thereafter Datum elevation 744.84 ft. msl

The maximum flood listed is known to be the greatest in at least — years (or since —).

Flood data for momentary peak discharges greater than 5,000 second-feet.

5618

YEAR	MONTH	DAY	GAGE HEIGHT (feet)	DISCHARGE (second-feet)	ANNUAL FLOODS ORDER (M)	ANNUAL FLOODS RECURRENCE INTERVAL (years)	PARTIAL DURATION SERIES ORDER (M)	PARTIAL DURATION SERIES RECURRENCE INTERVAL (years)	REFERENCE
1921	Dec.	24	13.20	11,800			29	0.86	
1922	Jan.	5	9.40	6,140			108	.23	
	Mar.	15	10.70	7,780			70	.36	
	Apr.	15	15.90	19,000	9	2.78	12	2.08	
	May	27	10.00	6,800			85	.29	
	June	17	12.00	9,700			43	.58	
1923	June	7	8.15	4,900	22	1.14			
1923	Dec.	6	9.90	6,690			90	.28	
		10	10.00	6,800			86	.29	
		14	10.10	6,940			83	.30	
		22	12.40	10,400			35	.71	
1924	Jan.	11	15.30	16,900			15	1.66	
	Feb.	20	10.50	7,500			74	.34	
	Mar.	5	10.10	6,940			84	.30	
		29	17.00	23,750	4	6.25	6	4.18	
	June	9	13.95	13,400			22	1.13	
1925	Mar.	19	8.20	4,900	21	1.19			
1926	Feb.	1	8.50	5,200			139	.18	
		19	9.08	5,810			119	.21	
		26	9.08	5,810			120	.21	
	Mar.	23	8.80	5,500			129	.19	
	Apr.	8	12.10	9,870	18	1.39	40	.62	
	Sept.	16	8.40	5,100			142	.18	
		24	12.00	9,700			44	.57	
1926	Oct.	6	10.20	7,080			81	.31	
		25	9.10	5,810			121	.21	
	Nov.	16	9.80	6,580			95	.26	
1927	Jan.	20	16.00	19,400	8	3.12	11	2.28	
	Mar.	21	15.50	17,600			14	1.79	
	May	19	10.70	7,780			71	.35	
1927	Dec.	1	12.90	11,200			32	.78	
		14	15.30	16,900	10	2.50	17	1.46	
1928	Feb.	8	9.60	6,360			100	.25	
		15	8.30	5,000			145	.17	
	Mar.	14	10.40	7,360			77	.32	
		30	9.30	6,030			110	.23	
	July	22	10.20	7,080			80	.31	On last sheet only.
				15,040	Mean annual flood for period 1921-45				

Notes:- Separate water years by drawing solid lines (——). Indicate breaks in period of record by double solid lines (══). Recurrence Interval = $\frac{N+1}{M}$.

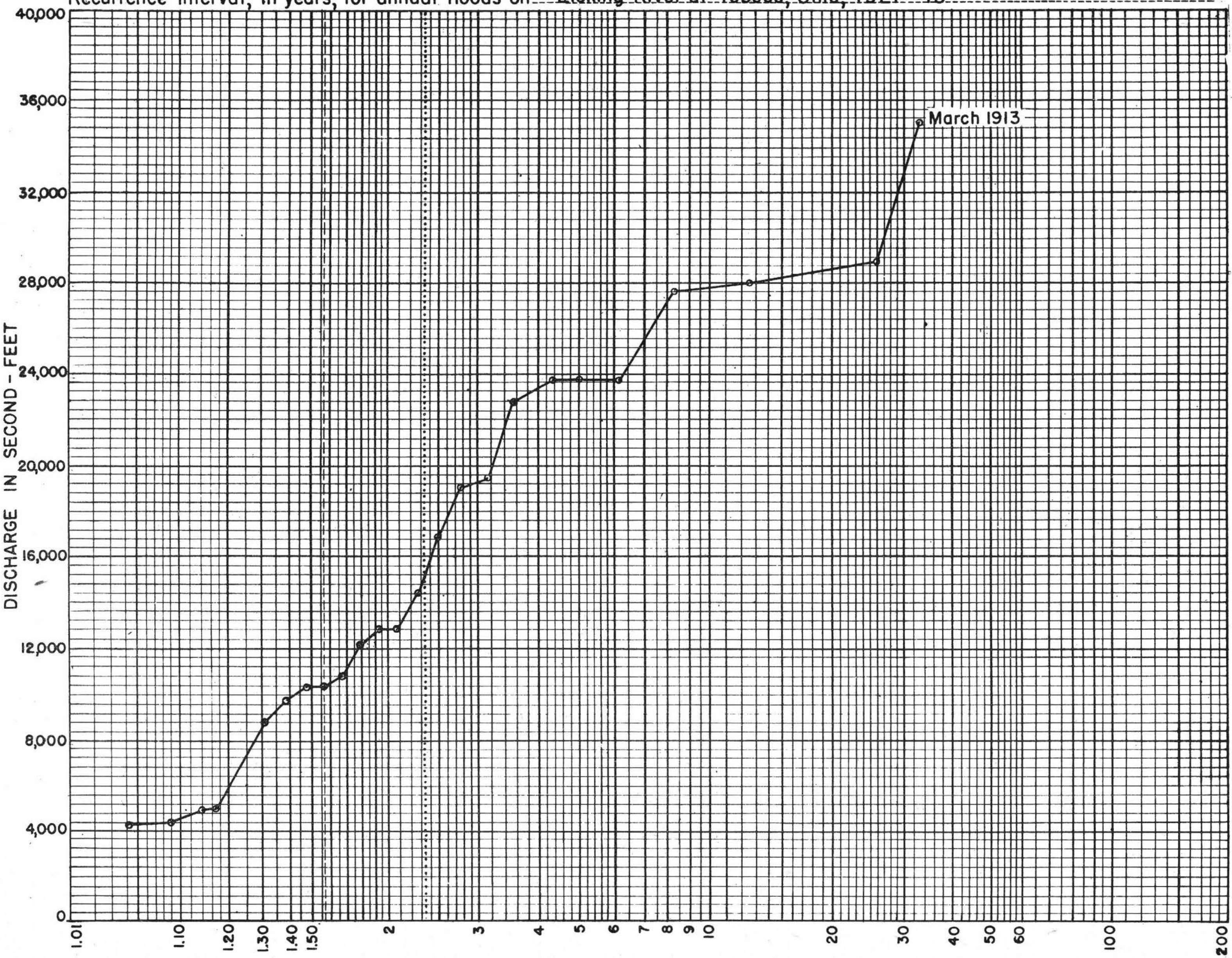
SHEET 1 OF 4. LISTED BY ________ DATE ________ CHECKED BY ________ DATE ________

9-179a
July 1946
Flood data plot
Experimental

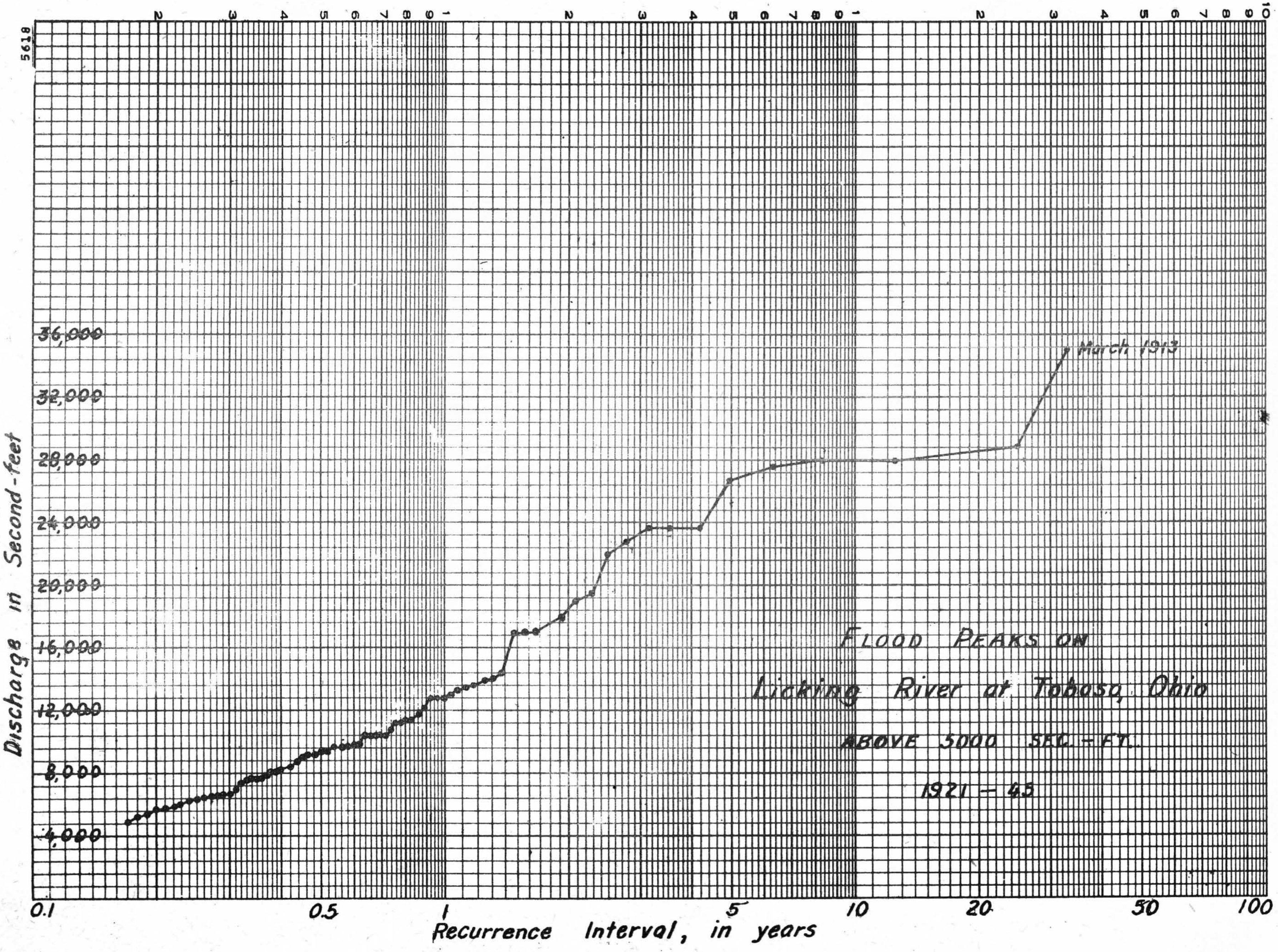
Department of the Interior—Geological Survey—Water Resources Branch

File Number ____________

Recurrence Interval, in years, for annual floods on Licking River at Toboso, Ohio, 1921-45



6


FLOOD PEAKS ON
Licking River at Toboso, Ohio
ABOVE 5000 SEC.-FT.
1921 – 45
March 1913
Discharge in Second-feet
36,000
32,000
28,000
24,000
20,000
16,000
12,000
8,000
4,000
Recurrence Interval, in years
0.1
0.5
1
5
10
20
50
100
5618