WATER RESOURCES RESEARCH

VOLUME 5 FEBRUARY 1969 NUMBER 1

PUBLISHED BY THE AMERICAN GEOPHYSICAL UNION

VOL. 5, NO. 1 WATER RESOURCES RESEARCH FEBRUARY 1969

Frequency Curves for Annual Flood Series with Some Zero Events or Incomplete Data

M. E. JENNINGS AND M. A. BENSON

U. S. Geological Survey, Washington, D. C. 20242

Abstract. In fitting a theoretical frequency distribution to a set of data, a problem arises if the series contains a number of zero values, as may occur in annual flood peak data for small, arid-region streams. The problem is twofold: first, commonly used distributions do not fit such a set of data; second, if a logarithmic transformation of the data is being used, logarithms of zero flows are not usable in a computation. To overcome the difficulties, a theorem of conditional probability is used. The probability of occurrence of a nonzero peak is combined with the conditional probability of exceeding a given flood magnitude, given that a nonzero peak has occurred. The method has been found useful also for fitting flood series in which information of peak annual floods below a specific stage is lacking.

The use of the log-Pearson Type III distribution has recently been recommended for fitting flood-frequency data by the *Water Resources Council* [1967]. In fitting the log-Pearson Type III flood distribution, or any other fitted distribution, the problem arises of how to handle stations with some zero events. Such events occur in arid or semiarid regions, particularly on small streams. One difficulty is that the logarithm of zero is minus infinity, and this makes a solution impossible. Another difficulty is that none of the commonly used theoretical probability distributions can fit a set of data part of which can be represented by a curve and part by a straight line of constant value. This report describes a method for solving the problem of zero peaks. It may also be used to obtain a frequency curve for stations where small peaks are not recorded; this may happen where a crest-stage gage has a lower limit of stage that is not exceeded during some years. The result of the method is a frequency curve with a continuous frequency segment and a discrete or mass probability segment. The technique is objective and, by comparison with the original data, appears to yield good results throughout the frequency range. *Beard* [1962], to correct for the effect of abnormal dry years, suggests ignoring lower flows and fitting a frequency curve to only the upper half of the ranked data. In the method proposed here, none of the data are ignored. The method is based on probability theory and may be explained by reference to Figure 1, which illustrates the sample space of flood events.

The conditional probability of an event Y, given that an event X has already occurred, as found in most texts on probability theory, is

$$P(Y/X) = P(YX)/P(X) \qquad (1)$$

Rearranging and noting from Figure 1 that $P(YX)$, the probability of the intersection of Y and X, is equal to $P(Y)$, equation 1 for the case of flood events is

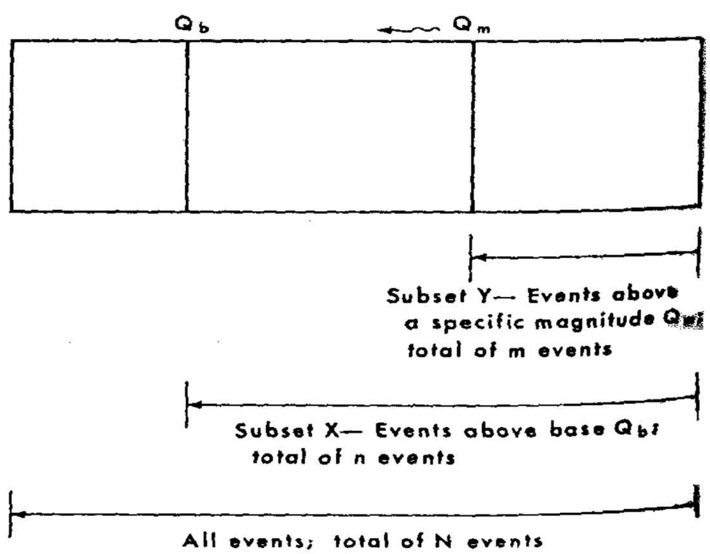

Fig. 1. Sample space of flood events.

TABLE 1*a*. Annual Peaks for La Brea Creek near Sisquoc, California

Flood Peak (Q_m), cfs	Order No., m	$\frac{m}{N+1}$
3320	1	0.042
1970	2	0.083
1600	3	0.125
1430	4	0.167
1360	5	0.208
785	6	0.250
513	7	0.292
275	8	0.333
227	9	0.375
191	10	0.417
178	11	0.458
155	12	0.500
12	13	0.542
1.9	14	0.583
9 zero peaks	15 to 23	

$$P(Y) = P(X) \cdot P(Y/X) \quad (2)$$

For flood events $P(X)$, the probability in any year of an event that exceeds Q_b, a base flood level equal to zero or a level above which flood magnitudes are recorded, is equal to n/N (see Figure 1 for meaning of m, n, N, and Q_m). In flood terminology, equation 2 can be written as

$$P\{\text{annual flood} > Q_m\} = \frac{n}{N} \cdot P\{\text{annual flood} > Q_m/\text{annual flood} > Q_b\} \quad (3)$$

The flood level Q_m, where $m = 1, 2 \ldots n$ is the order number of the arrayed flood magnitudes, is reduced in steps toward Q_b and eventually equals Q_b as selected flood levels and their associated probabilities are determined. The probability $P\{\text{annual flood} > Q_m\}$ or plotting position of Q_m is from Figure 1 equal to m/N. The

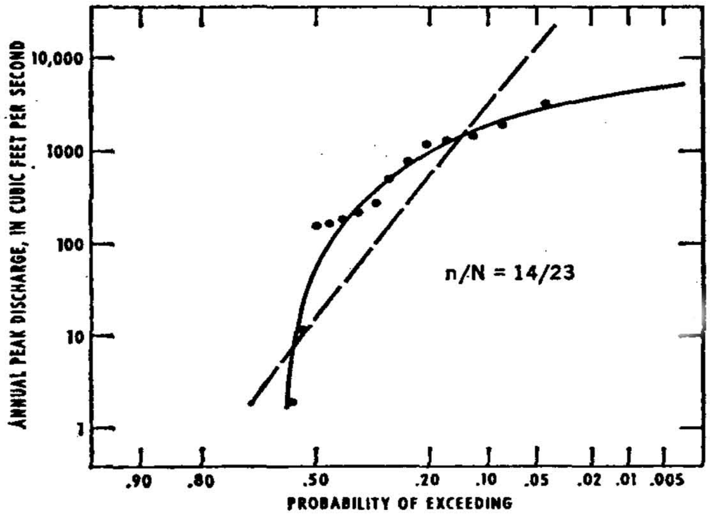

Fig. 2. Log-Pearson Type III frequency curve for La Brea Creek near Sisquoc, Calif. The dashed curve is the frequency curve obtained when 0.1 cfs is added to all flood peaks.

ratio m/N is based on observed events. Numerically, it is close to the plotting position formula $m/(N + 1)$, which represents the expected probability of ranked events [*Langbein*, 1960], which is recommended for use in the graphical plotting of individual flood events. The latter formula may be used to assign probabilities to the individual events in a flood series. After plotting these probabilities against the flood magnitudes, a graphical frequency curve may then be drawn based on these events.

If, however, it is desired to fit a theoretical distribution such as the log-Pearson Type III distribution to the flood series, the conditional probabilities in equation 3 may be defined by fitting a log-Pearson Type III distribution to events greater than Q_b. These probabilities multiplied by n/N are the probabilities of annual floods greater than Q_m.

Table 1 lists the annual peaks for La Brea Creek near Sisquoc, California. Because only 14

TABLE 1*b*. Log-Pearson Type III Results for Nonzero Peaks

Conditional Probability, $P(Y/X)$	0.90	0.50	0.20	0.10	0.05	0.02	0.005
Q_m in cfs	16.1	477	1720	2680	3530	4390	5230
$P\{\text{annual flood} > Q_m\}$ or $P(Y)$	0.548	0.304	0.122	0.061	0.030	0.012	0.0030

of the 23 annual peaks are greater than zero, $n/N = 14/23$. The conditional probabilities as shown in equation 3 and corresponding discharges Q_m shown on the first two lines of Table 1*b* were taken directly from the computer output sheet and are the results of a log-Pearson Type III fit using only non-zero events. These probabilities are multiplied by 14/23 to obtain the probabilities of an annual flood being greater than Q_m, shown on line 3 of Table 1*b*.

To compare the fitted curve with the original data, the observed peaks listed in Table 1*a* are arrayed by magnitude and plotted in Figure

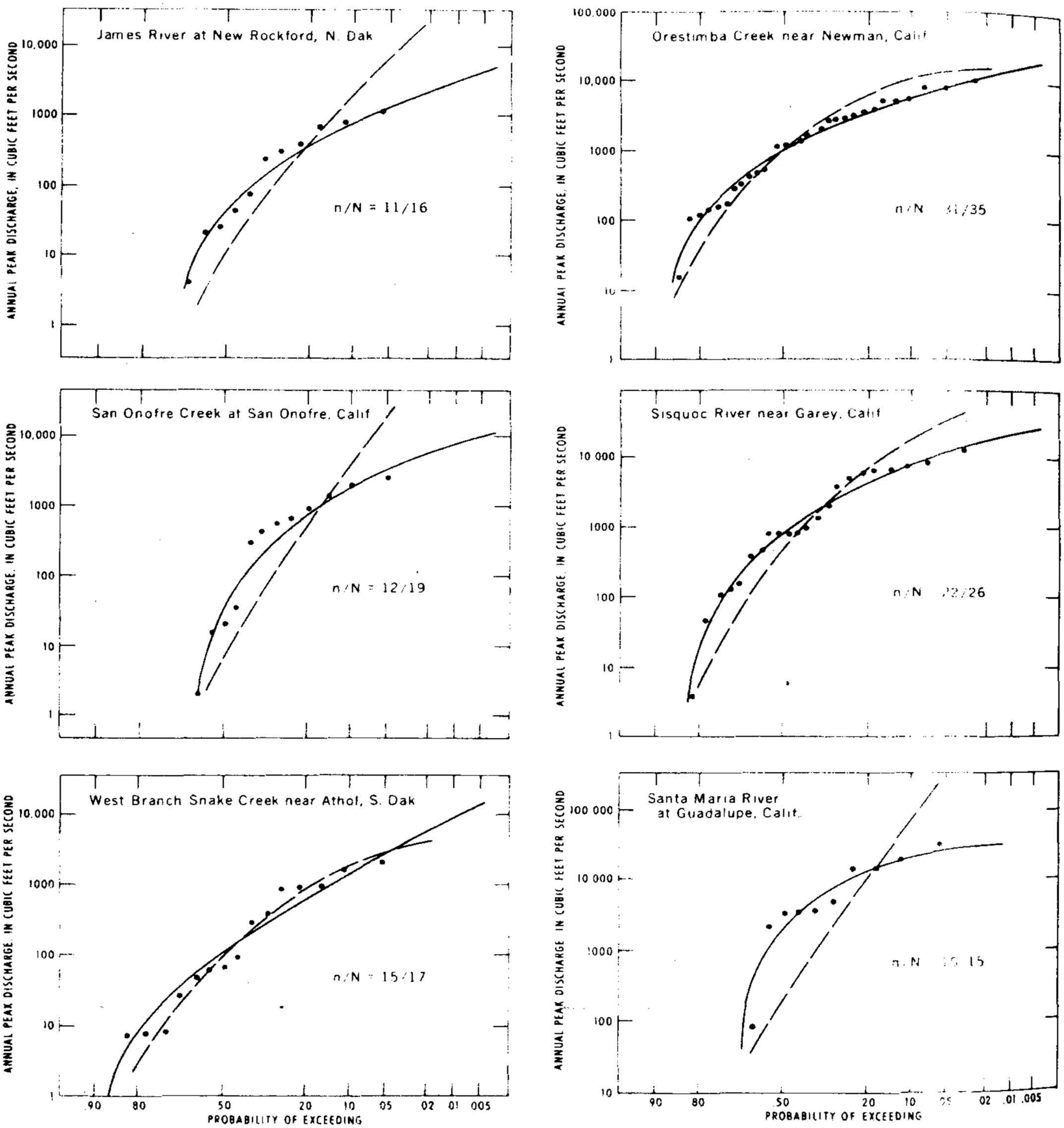

Fig. 3. Log-Pearson Type III frequency curves for the case of Q_0 equal zero. The dashed curves are the frequency curves obtained when 0.1 cfs is added to all flood peaks.

TABLE 2*a*. Annual Peaks for Antelope Creek Tributary No. 2 near Harlowton, Montana

	Flood Peak (Q_m), cfs	Order No., m	$\frac{m}{N+1}$
	3230	1	0.083
	922	2	0.167
	820	3	0.250
	190	4	0.334
a	130	5	0.417
	80	6	0.500
f	<50	...	
	35	...	
f	<25	...	
f	<25	...	
a	20		

a—about f—below recording level of gage

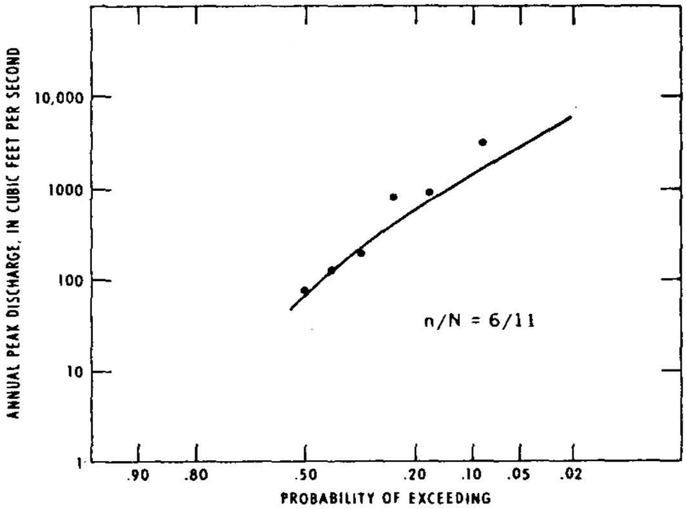

Fig. 4. Example of log-Pearson Type III frequency curve for Q_b greater than zero.

2 using the plotting position formula $m/(N+1)$, where m is the order number. The solid line shown is the fitted curve.

Annual flood data from six additional gaging stations with different numbers of zero peaks were used to test the method. The results, shown in Figure 3, are considered good.

Another way of treating zero events [*Subcommittee on Hydrology,* 1966] is to add a small discharge to all flood events and thus fit a log-Pearson Type III curve to the data. This method was tried for the stations analyzed by adding 0.1 cfs to all flood events. The results are shown as dashed curves in Figures 2 and 3 and, in general, indicate poorer fitting than by the method proposed in this paper.

An example of a flood series with Q_b at a level above zero is presented using data for a crest-stage station in Montana. The annual peak data and resulting frequency curve for Antelope Creek tributary No. 2 near Harlowton, Montana, are given in Tables 2*a* and 2*b* and Figure 4. For this computation, Q_b was set at 50 cfs, since annual floods below this level were not recorded.

In summary, the proposed procedure for preparing a frequency curve using the log-Pearson Type III method for an annual flood series having some zero events or incomplete data is outlined in the following steps:

1. List annual flood events above Q_b, where Q_b is zero or something higher.

2. Fit a log-Pearson Type III curve to these events, either by computer or by manual calculations, and thus obtain discharge values corresponding to conditional probabilities throughout the defined range.

3. Calculate n/N as the proportion of peaks above Q_b, and calculate $P\{\text{annual flood} > Q_m\}$ using equation 3. For the case of Q_b equal to zero, plot $P\{\text{annual flood} > Q_m\}$ and corresponding Q_m discharges on log probability paper

TABLE 2*b*. Log-Pearson Type III Results for Peaks Greater than 50 cfs

Conditional Probability, $P(Y/X)$	0.90	0.80	0.50	0.20	0.10	0.04
Q_m in cfs	72	123	378	1317	2620	5720
$P\{\text{annual flood} > Q_m\}$ or $P(Y)$	0.490	0.436	0.272	0.109	0.055	0.022

and draw a smooth curve. For the case of Q_b equal to some value greater than zero, plot n/N at Q_b and $P\{\text{annual flood} > Q_m\}$ corresponding to Q_m discharges on log probability paper and draw a smooth curve terminating at Q_b and ignoring discharges that may be less Q_b. The purpose of drawing curves in the case of either the zero or the nonzero base is only to permit the interpolating of flood magnitudes corresponding to desired probabilities or recurrence intervals. This interpolation could be done analytically if desired, but the graphical interpolation is simpler.

4. Compare the fitted curve with the observed data as shown in Figures 2–4.

REFERENCES

Beard, L. R., Statistical methods in hydrology, *U. S. Army Corps of Engineers, Sacramento, California*, 62 pp. and 2 appendices, 1962.

Langbein, W. B., Plotting positions in frequency analysis, *in* Dalrymple, T., Flood-frequency analysis, *U. S. Geol. Surv. Water-Supply Paper 1543-A*, 48–51, 1960.

Subcommittee on Hydrology, Methods of flow frequency analysis, *Bull. 13, Inter-Agency Comm. on Water Resources*, Government Printing Office, Washington, D. C., 42 pp., 1966.

Water Resources Council, A uniform technique for determining flood flow frequencies, *Bull. 15, Hydrol. Comm., Water Resources Council*, 1025 Vermont Ave., N. W., Washington, D. C. 20005, 1967.

(Manuscript received October 14, 1968.)