A Comparison of Methods Used In Flood-Frequency Studies for Coastal Basins in California

GEOLOGICAL SURVEY WATER-SUPPLY PAPER 1580-E

Prepared in cooperation with the California Department of Water Resources

A Comparison of Methods Used In Flood-Frequency Studies for Coastal Basins in California

By R. W. CRUFF *and* S. E. RANTZ

FLOOD HYDROLOGY

GEOLOGICAL SURVEY WATER-SUPPLY PAPER 1580-E

Prepared in cooperation with the
California Department of Water Resources

UNITED STATES GOVERNMENT PRINTING OFFICE, WASHINGTON : 1965

UNITED STATES DEPARTMENT OF THE INTERIOR

STEWART L. UDALL, *Secretary*

GEOLOGICAL SURVEY

Thomas B. Nolan, *Director*

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402 - Price 70 cents (paper cover)

CONTENTS

ILLUSTRATIONS

[All plates are in pocket]

TABLES

FLOOD HYDROLOGY

A COMPARISON OF METHODS USED IN FLOOD-FREQUENCY STUDIES FOR COASTAL BASINS IN CALIFORNIA

By R. W. Cruff and S. E. Rantz

ABSTRACT

This study compares the results of regional flood-frequency studies made by several methods and appraises the relative reliability of these methods. The areas selected for study were the subhumid San Diego area in southwestern California and the humid coastal area in northwestern California. The following six methods of analysis were applied to each region: Index-flood method, multiple correlation, logarithmic normal distribution, extreme-value probability distribution (Gumbel method), Pearson type III distribution, and gamma distribution. The last four methods named involved not only the computation of the statistics appropriate to the distributions, but also the relating of these statistics to basin and climatologic characteristics. On the basis of an empirical, nonstatistical test, the following conclusions were reached:

1. All methods of analysis give better results in a humid region than in a subhumid region because streamflow is less variable in a humid region.
2. If historical data, either qualitative or quantitative, are available concerning the magnitude of floods that occurred in the years prior to the collection of streamflow records, the multiple-correlation method of analysis is preferred. Only this method and the index-flood method benefit from the historical data, which, in effect, extend the time base of the analysis. The multiple-correlation method is superior to the index-flood method because it has a far more rational basis and in addition gives better results.
3. Where the peak-discharge data are limited entirely to the period during which streamflow records were collected (no historical data available), a method based on the distribution of the array of peak flows is preferred because of its greater objectivity. Of the four distributions tested, the Pearson type III is the most desirable. It is more flexible than the other three and will generally fit the peak-discharge data best.

Although this comparison study of flood-frequency methods was based on small samples from only one part of the United States, the results and conclusions appear to be meaningful because they can be explained rationally.

INTRODUCTION

The principle of analyzing flood magnitudes on a probability basis is almost universally accepted because its use permits economic considerations, as well as hydrologic factors, to govern the planning and design of projects that are susceptible to flood damage. There is no universal acceptance, however, of any single method of making the flood-frequency analysis. Usually, a basic objective of the analysis is to derive flood magnitude-frequency relations which may be used at a site for estimating the magnitude of rare or unusual flood events, such as the peak discharge that has an average probability of being exceeded only once in 100 years (Q_{100}) or of being exceeded only once in 50 years (Q_{50}). Because our records of flood discharges are generally short—less than 30 years, on the average—extrapolation by some means is required to estimate the magnitudes of unusual floods. However, the magnitudes so determined depend, to a large degree, on the method of frequency analysis that governed the extrapolation. Therefore, it is not unusual for independent workers, using the same short streamflow records but different methods of analysis, to obtain widely differing values of discharge corresponding to Q_{50} or Q_{100}. Furthermore, because of the element of uncertainty that characterizes any extrapolation, it is seldom possible to decide which of the derived discharges are the most accurate or which method of analysis is the most reliable.

For a detailed discussion of the many methods of deriving flood magnitude-frequency relations, the reader is referred to reports by Jarvis and others (1936) and Benson (1962). Several of the methods described in those reports are no longer in favor, and others have had varying degrees of popular acceptance over the years. At present (1964), the methods most commonly used in this country are the four listed below. The agency or agencies shown in parentheses are the chief proponents of the methods.

1. Index-flood method (U.S. Geol. Survey).
2. Multiple-correlation (U.S. Bur. of Public Roads and U.S. Geol. Survey).
3. Logarithmic normal distribution (U.S. Army, Corps of Engineers).
4. Extreme-value probability distribution or Gumbel method (U.S. Weather Bur.).

These four methods and two additional ones—the Pearson type III distribution and the gamma distribution—were used in this study. The Pearson type III distribution was included because the authors feel that this method is likely to regain the popularity it once had in probability studies of peak discharge. The gamma distribution was included because it is increasingly being used, both in the United States and abroad, for studying the probability of occurrence of hydrologic events, including peak discharges.

PURPOSE AND SCOPE

The purpose of this study was to compare the results obtained by applying each of the six methods of flood-frequency analysis to identical sets of peak-discharge data. Specifically, it was the values of Q_{50} and Q_{100}, computed by each method, that were compared. One comparison was made for streams in the San Diego area in south coastal California, and another was made for streams in north coastal California. (See fig. 1.) There were several reasons for selecting these two areas. First, the comparison is given broader scope by using two areas that differ greatly in the amount of precipitation they receive; the San Diego area is subhumid, whereas the north coastal area is very humid. Second, flood-frequency studies have recently been published for each area, and the hydrologic factors needed in the analyses were therefore readily available. The San Diego area study, published by the California Department of Water Resources (1963), analyzed the data for 18 stream-gaging stations by two methods—the index-flood method and multiple correlation. Of the 18 peak-discharge records, 2 were longer than 35 years, 11 were between 20 and 35 years, and 5 were shorter than 20 years. The study for north coastal California published by the U.S. Geological Survey (Rantz, 1964) analyzed the data for 27 stream-gaging stations by the index-flood method. Of the 27 peak-discharge records, 4 were longer than 30 years, 9 were between 10 and 30 years, and 14 were shorter than 10 years.

The principal reason for selecting the two areas lay in the fact that, although the discharge records for the streams were relatively short, the magnitudes of the greatest and second greatest flood peaks in the past 100 years were generally known within reasonable limits of accuracy. With this knowledge it was possible to appraise the results obtained by each of the six methods of analysis and to draw conclusions concerning the relative reliability of the methods.

Many competent statisticians, using long-term discharge records and rigorous statistical treatment, have made comparison studies of the various methods of analyzing flood magnitude-frequency relations without arriving at definitive conclusions concerning the superiority of any one method. Hence the plethora of methods of analysis from which the practicing engineer must choose before making a flood-frequency study. In view of the many uncertainties involved in a comparison study of methods of analysis, the authors have eschewed any rigorous statistical approach, such as an analysis of variance or a determination of confidence limits. They have adopted, instead, a simple pragmatic approach in which the basis of comparison is the relative ability of the various methods to reproduce Q_{50} and Q_{100} at each of the study sites in the two California areas. This approach

770-401 O—65——2

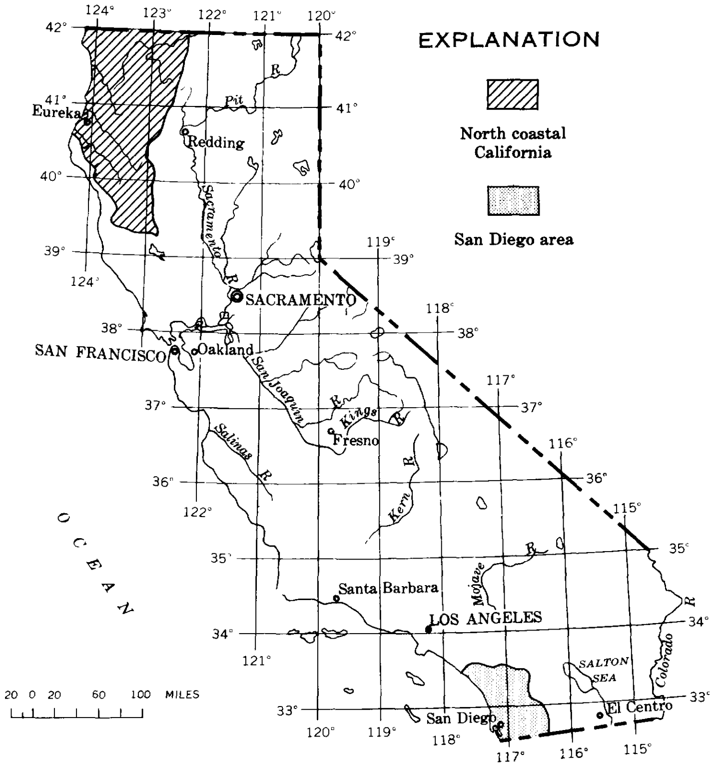


FIGURE 1.—Location of areas studied.

has its statistical shortcomings but is not unreasonable. After all, the objective of all methods is to fit some relation to the available peak-discharge data, and because it is the magnitudes of the infrequent floodflows that are usually sought, the method that best fits the data for the infrequent floods is presumably the most reliable method for use. Admittedly, the conclusions reached in this report through this pragmatic approach primarily represent impressions gained during the course of the study, but the findings appear to be meaningful.

ACKNOWLEDGMENTS

This study was made under the terms of a cooperative agreement between the U.S. Geological Survey and the California Department of Water Resources. The report was prepared by the Surface Water Branch of the Geological Survey under the general supervision of Walter Hofmann, district engineer. E. D. Cobb, hydraulic engineer, assisted in the computation and preparation of the data.

The authors are grateful for the many helpful suggestions of D. R. Dawdy, J. R. Crippen, M. A. Benson, and H. C. Riggs, hydraulic en-

gineers of the Geological Survey, who reviewed the manuscript. However, the subject treated here is not one on which unanimity of opinion can be expected. The opinions and assertions in this report are those of the authors and do not, in all particulars, reflect the views of their coworkers in the Geological Survey.

STATISTICAL EQUATIONS USED IN THIS STUDY

We assume that the reader has some knowledge of elementary statistics and is familiar with common statistical nomenclature and with the equations for computing such elemental items as mean, standard deviation, correlation coefficient, and linear regression equation. These terms and equations can be found in any standard statistics text (for example, Ezekiel and Fox, 1959). All other statistical equations that are used in this report are explained where they first appear and are thereafter referred to by number as equation 1, equation 2, and so on.

DESCRIPTION OF THE METHODS OF ANALYSIS USED IN THIS STUDY

In this section of the report, the six methods of analysis used in this study are briefly described. In all six methods the streamflow data analyzed consist of the momentary maximum discharge of each year of record. Gaging stations whose peak discharges are seriously affected by manmade storage or diversion are not used. All six methods attempt to make the most efficient use of the data available. The index-flood and multiple-correlation methods are to some extent empirical because the graphical curve-fitting procedures that are basic to both methods require a certain amount of subjectivity on the part of the analyst. These methods have the advantage of permitting the analyst to use whatever qualitative or historical information is available for extrapolating the flood-frequency relation. The other four methods are empirical in the sense that a type of distribution—logarithmic normal, extreme-value probability, Pearson type III, or gamma—is arbitrarily selected for use. One may theorize concerning the probability distribution that describes the occurrence of flood events, but the lack of agreement among hydrologists indicates that the "true" distribution is not known. Once a distribution is selected for use, however, the analysis becomes strictly objective, and the extrapolation is automatically made from a mathematical determination of the statistics—mean, standard deviation, and coefficient of skew—of the station data. Qualitative or historical data cannot be assessed, however, when a rigorous statistical solution is made. On occasion, this type of information is used to define or modify the high-water end of the computed flood-frequency relation; but when this information is so used, the statistical distribution no longer controls the all-important high-water end of the flood-frequency relation. As in the empirical index-flood and multiple-correlation methods, the results obtained will depend to a large degree on the judgment of the analyst in interpreting the historical information and assigning probabilities to the peak discharges. In this report, an objective mathematical treatment was used for the four methods involving standard statistical distributions, with no consideration given to qualitative or historical information.

Regardless of the method of analysis used, it is customary to develop flood magnitude-frequency relations that are applicable to an entire region rather than to a single gaging station. Because the flood series for a single station is a short random sample, it may not be representative of the long-term distribution of flood events at the gaging station. Combining records for all stations in a hydrologically homogenous area tends to reduce the sampling error associated with a nonrepresentative sample. Another advantage of the regional flood-frequency relation is that it can be applied to ungaged sites in the region. The boundaries of a homogeneous region must be rationally delineated from a knowledge of the hydrology of the region; commonly, these boundaries will coincide with the boundaries of physiographic sections delineated by Fenneman (1931, 1938) or with the boundaries of regions delineated on soil classification maps published by the U.S. Soil Conservation Service. In making a regional study, a common base period of years is usually used for each gaging station. This base period is generally the period of record for the older gaging stations in the region, and the shorter records are extended by correlation procedures to cover the base period.

Generally, the term "regionalization," as used by engineers engaged in flood-frequency analysis, refers not only to the delineation of the boundaries of hydrologically homogeneous regions, but also to the establishment of relations between pertinent characteristics of the flood-frequency curve and basin or climatologic parameters within the homogeneous region. For example, values of the mean annual flood are said to be regionalized when a relation is found between mean annual flood and size of drainage area in the region being studied. The terms "regionalization," or "regionalized" are used in this broad sense in this report.

INDEX-FLOOD METHOD

The index-flood method (Dalrymple, 1960) has been the standard U.S. Geological Survey method of flood-frequency analysis for the past 15 years. A step-by-step outline of the procedure follows:

1. Peak-discharge data within the base period are tabulated for each station with 10 or more years of record.
2. Historical or qualitative information is noted for each station. As an example, suppose the base period is 1930–60. Information of the following types may be available:

   "Flood of 1955 is of approximately the same magnitude as that of 1862, the greatest previously known." Or, "Prior to 1955, the flood of 1932 was the greatest known since the flood of 1893."
3. Peak discharges needed for each short-term station to complete the record for all years of the base period are computed from a regression line or equation. The regression is obtained by graphically correlating concurrent peak discharges for the short-term station and a nearby long-term station.
4. The peak discharges at each station are ranked in order of magnitude, starting with 1 for the greatest discharge, 2 for the second largest discharge, and so on.
5. The recurrence interval for each *observed* peak discharge is computed. This is not done for the peak discharges computed by regression equation in step 3. The only purpose of the computed peak discharges is to provide a basis on which to estimate the recurrence interval for the observed peak discharges. The formula used to compute recurrence interval is:

$$RI=\frac{n+1}{m}, \tag{1}$$

   where

   RI is the recurrence interval in years,

   n is the years of record, and

   m is the order of magnitude of an annual peak discharge.
6. Recurrence intervals are adjusted, where appropriate, on the basis of historical or qualitative information. In the example in step 2, n for the base period is 31 years. The flood of 1955, being the greatest of record during the base period, would normally have its recurrence interval computed as 32 years. However, in the past 99 years, a flood of this magnitude occurred twice—in 1862 and in 1955—but has not been exceeded. Therefore, a peak discharge equivalent to that of 1955 has orders of magnitude 1 and 2 in 99 years, giving it recurrence intervals of 50 and 100 years, instead of the single recurrence interval of 32 years that was originally computed. From the statement regarding the flood of 1893, we know that the peak discharge of 1932 was not merely the second largest in 31 years, but was the third largest in at least 68 years, and it therefore has a recurrence interval of 23 years rather than the 16 years originally computed.

(We assume in this example that the flood of 1893 is greater than that of 1932 but that its peak discharge is unknown.)

7. For each station, the recurrence interval is plotted in relation to peak discharge on extreme-value probability graph paper. A straight line or gentle curve is fitted by eye to the plotted points.
8. The mean annual flood ($Q_{2.33}$), defined as the discharge corresponding to a recurrence interval of 2.33 years on the graph described in the preceding step, is selected for each station.
9. The peak discharge corresponding to a 10-year recurrence interval (Q_{10}) on the graphs described in step 7 is selected for each station.
10. The comparative slope of the individual curves between Q_{10} and $Q_{2.33}$, or the ratio of Q_{10} to $Q_{2.33}$, is computed for each station.
11. From a knowledge of the hydrology of the region being studied, areas are delineated that are expected to have similar ratios of Q_{10} to $Q_{2.33}$. Commonly, the boundaries of the areas will be governed by: (a) physiography (similar ratios would be expected where drainage basins have similar shape; where precipitation may occur as either rain or snow, high-altitude basins would tend to have one ratio and low-altitude basins another) and (b) mean annual precipitation (generally, the more humid the area, the smaller the ratio).
12. A homogeneity test (Dalrymple, 1960, p. 38–39) is made of the ratios of Q_{10} to $Q_{2.33}$. On the basis of this test, the boundaries of the areas selected in step 11 may be adjusted, but a rational explanation for any changes that are made is desirable.
13. At each station all recorded discharges are divided by $Q_{2.33}$; so these discharges are expressed as dimensionless ratios.
14. The median dimensionless discharge ratio for each recurrence interval is determined for each group of stations in the homogenous areas selected in step 11 or 12.
15. For each homogeneous area, median dimensionless discharge ratios are plotted in relation to recurrence interval on extreme-value probability paper, and a straight line or curve is fitted by eye to the plotted points.
16. The procedures described in steps 1–15 apply only to those stations with 10 or more years of peak-discharge record. Peak discharge data are next tabulated for those stations with 5–9 years of record within the base period.
17. Concurrent peak discharges for each of these short-term stations and for a nearby longer term station are correlated graphically. $Q_{2.33}$ for the short-term station is then determined from the regression line, it being the discharge corresponding to $Q_{2.33}$ for the longer term station.

18. From a knowledge of the hydrology of the region being studied, areas of probable homogeneity with regard to the mean annual flood are delineated. Generally, this homogeneity refers to a similarity in infiltration characteristics. As mentioned earlier, the boundaries of these homogeneous areas will often coincide with the boundaries of physiographic sections delineated by Fenneman (1931, 1938) or with the boundaries shown on Soil Conservation Service soil classification maps. These boundaries, however, may or may not coincide with those delineated in step 11 or 12.
19. Within each of the areas from step 18, $Q_{2.33}$ for both long- and short-term stations (values obtained from steps 8 and 17) is correlated graphically with drainage area and with other significant parameters such as mean annual rainfall, areas of lakes and ponds, main-channel slope, and mean basin altitude.

The correlation graph from step 19 and the dimensionless flood-frequency curve from step 15 are the end products of this analysis. Used for the appropriate region, the first graph provides a means of estimating $Q_{2.33}$ from basin and climatologic parameters; the second graph is a regional flood-frequency curve in which discharges are expressed as a ratio to $Q_{2.33}$.

MULTIPLE CORRELATION

The multiple-correlation method of flood-frequency analysis is becoming increasingly popular in the U.S. Geological Survey. In performing a flood-frequency analysis by this method, a region of probable hydrologic homogeneity is first selected as previously described. For each gaging station in the region with 10 or more years of peak-discharge record within the base period of years an individual flood-frequency curve is drawn by following steps 1–7 that are described in the preceding section entitled "Index-flood method." After the station flood-frequency curves have been prepared, discharges are read at selected recurrence intervals, such as 2.33 years ($Q_{2.33}$), 5 years (Q_5), 10 years (Q_{10}), 20 years (Q_{20}), 50 years (Q_{50}), and 100 years (Q_{100}). Each set of discharges is then correlated with various basin and climatologic parameters, using a regression equation of the form:

$$Q_T = aB^bC^cD^d \ldots\ldots, \qquad (2)$$

where

Q_T is the discharge corresponding to a recurrence interval of T years,

a, b, c, d . . . are the constants, and

B, C, D . . . are the basin and climatologic parameters.

The basin and climatologic parameters that are considered include drainage area, mean annual precipitation, area of lakes and ponds, land slope, main-channel slope, mean basin altitude, a shape factor, and others. The constants in the equation are computed by least squares, and statistical tests are made to eliminate from the equation those parameters that have little or no significance. From the final equations for discharges corresponding to selected recurrence intervals, a flood-frequency curve can be constructed for any site in the region, whether gaged or ungaged, once the values of the significant parameters are determined.

The U.S. Bureau of Public Roads uses a variation of this method for analyzing floods from small drainage areas. The Bureau method (Potter, 1961) involves a graphical correlation between discharges corresponding to a selected recurrence interval and three hydrologic parameters—drainage area, a precipitation index, and a topographic index. Standard curves have been prepared for several regions in the country.

LOGARITHMIC NORMAL DISTRIBUTION

In making flood-frequency studies, the U.S. Army, Corps of Engineers, bases its analysis on the premise that the logarithms of annual peak discharges are normally distributed. Beard (1962) prepared a detailed description of the method used by the Corps of Engineers; a brief step-by-step résumé of the method is given below. Only those stations with 10 or more years of peak-discharge record within the base period of years are used in the analysis.

1. Logarithms of the peak-discharge data within the base period are tabulated for each station in a region of probable hydrologic homogeneity.
2. The mean and the standard deviation of the array of data for each station are computed.
3. The mean and the standard deviation for each short-term station are adjusted to cover the complete base period. This is done by first computing a linear correlation for concurrent peak discharges at the short-term station and at a nearby long-term station. The adjustment of the mean and the standard deviation is then made by means of the following equations:

$$S_{1b}-S_{1a}=(S_{2b}-S_{2a})(\overline{R})^2(S_{1a}/S_{2a}) \tag{3}$$

and

$$M_{1b}-M_{1a}=(M_{2b}-M_{2a})(\overline{R})^2(S_{1b}/S_{2b}), \tag{4}$$

where

M is the mean of the logarithms of the peak discharges,

S is the standard deviation of the logarithms of the peak discharges,

$\bar{R}$ is the coefficient of correlation adjusted for lost degrees of freedom,

$_1$ is the short-term station,

$_2$ is the long-term station,

$_a$ is the short-term period, and

$_b$ is the base period.

These formulas have some statistical shortcomings, but they tend to give an unbiased estimate of the all-important long-term standard deviation.

4. To regionalize the statistics of the logarithmic normal distribution, the base-period mean (M) and base-period standard deviation (S) of the logarithms of peak discharge are each correlated with basin and climatologic parameters in the homogeneous region. (Remember that in a logarithmic normal distribution the antilog of the mean of the logarithms of the peak discharges corresponds to the geometric mean or to the median of the natural values of these discharges and not to their arithmetic mean.)
5. From the relations obtained in step 4, a flood-frequency curve can be constructed for any site in the region, whether gaged or ungaged, by use of the equation:

$$Q_T = M + K_T S, \quad (5)$$

where

Q_T is the logarithm of the discharge corresponding to a recurrence interval of T years,

M is the mean of the logarithms of annual peak discharges,

S is the standard deviation of the logarithms of annual peak discharges, and

K_T is a characteristic of the normal distribution; for the purpose of this report it may be defined as a coefficient corresponding to a recurrence interval of T years. (The following table gives values of K corresponding to selected values of recurrence interval, RI.)

RI (years)	K	RI (years)	K
2	0.00	20	1.64
5	.84	50	2.05
10	1.28	100	2.33

6. If plotted on logarithmic normal probability graph paper, the computed flood-frequency relation will be a straight line.

770-401 O—65——3

EXTREME-VALUE PROBABILITY DISTRIBUTION OR GUMBEL METHOD

About 20 years ago E. J. Gumbel began advocating the use of an extreme-value probability distribution for analyzing the magnitude-frequency relation of annual peak discharges, and his method still enjoys great favor among hydrologist-statisticians. Although a more recent publication by Gumbel (1958, p. 236, 272) describes three basic distributions of extreme values which may be used for flood studies, the distribution he originally advocated is still the most widely used. The U.S. Weather Bureau is one of the chief proponents of the Gumbel method and uses it for both precipitation-frequency and flood-frequency studies at individual sites.

To apply the Gumbel method, an area of probable hydrologic homogeneity is first selected. The base-period mean and the standard deviation of annual peak discharges are then computed for each gaging station in the region with 10 or more years of peak-discharge record within the base period by following steps 1–3 that are described in the preceding section entitled "Logarithmic normal distribution." The only difference is that in the Gumbel method natural values of peak discharge are used, and not their logarithms. The base period mean (M) and standard deviation (S) are then regionalized by correlation with basin and climatologic parameters in the homogeneous region. From these regional relations, a flood-frequency curve can be constructed for any site in the region, whether gaged or ungaged, by use of the formula:

$$Q_T = M + K'_T S, \qquad (6)$$

where

Q_T is the discharge corresponding to a recurrence interval of T years,
M is the mean of the peak discharges,
S is the standard deviation of the peak discharges, and
K'_T is a characteristic of the extreme-value probability distribution; for the purpose of this report it may be defined as a coefficient corresponding to a recurrence interval of T years. (The following table gives values of K' corresponding to selected values of recurrence interval, RI.)

RI (years)	K'	RI (years)	K'
2.33	0	20	1.87
5	.72	50	2.59
10	1.30	100	3.14

If plotted on extreme-value probability graph paper with arithmetic ordinate, the computed flood-frequency relation will be a straight line.

PEARSON TYPE III DISTRIBUTION

The Pearson type III distribution, or such variations of it as are expressed in the Hazen and Foster methods (Jarvis, 1936), was at one time widely used in probability studies of peak discharge. Its popularity for this purpose has declined in the past 20 years, but there is now renewed interest in its use. The Corps of Engineers uses the Pearson type III distribution in probability studies of flood volume, in which the annual flood volumes for various durations are expressed as logarithms. In this study natural values of annual peak discharges are used.

The Pearson type III distribution is more flexible than either the logarithmic normal or Gumbel distributions and can be more closely fitted to the data because it is defined not only by the mean and the standard deviation of the array of flood peaks but also by the coefficient of skew of the array. A discussion of the Pearson III distribution is given by Elderton (1953). A step-by-step description of its application in a flood-frequency study follows. Only those stations with 10 or more years of peak-discharge record are used in the analysis.

1. Peak-discharge data within the base period are tabulated for each station in a region of probable hydrologic homogeneity.
2. The mean and the standard deviation of the array of data for each station are computed.
3. The mean and the standard deviation for each short-term station are adjusted to cover the base period. This is done by first computing a linear correlation for concurrent peak discharges at a short-term station and at a nearby long-term station. The adjustments are then made by applying equations 3 and 4 used in step 3 of the section entitled "Logarithmic normal distribution." (Note that natural values of the discharges, and not their logarithms, are now used in the two equations.)
4. Before computing coefficients of skew, the individual values of peak discharge at the short-term stations that are needed to complete the record for all years of the base period are next computed by means of the equation:

$$X_1 - M_1 = (X_2 - M_2)(S_1/S_2), \tag{7}$$

where

X_1 and X_2 are the peak discharges for any given year at stations 1 and 2, the short- and long-term stations, respectively,

M_1 and M_2 are the mean values of annual peak discharge for concurrent periods at stations 1 and 2, respectively, and

S_1 and S_2 are the standard deviations of the annual peak discharges for concurrent periods at stations 1 and 2, respectively.

Equation 7 is used rather than a regression equation to minimize bias in the standard deviation of the final array of observed and computed peak discharges.

5. The coefficient of skew is computed for each station from the equation:

$$g=\frac{(N^2)(\Sigma X^3)-(3N)(\Sigma X)(\Sigma X^2)+2(\Sigma X)^3}{N(N-1)(N-2)(S)^3}, \tag{8}$$

where

g is the coefficient of skew,
N is the number of years of record (base period),
X is the magnitude of a peak discharge, in cubic feet per second, and
S is the standard deviation, in cubic feet per second.

All peak discharges, both observed and computed by equation 7, are used in determining coefficients of skew.

6. The base-period mean (M), standard deviation (S), and coefficient of skew (g) are regionalized by correlation with basin and climatologic parameters in the homogeneous region.
7. From the relations obtained in step 6, a flood-frequency curve can be constructed for any site in the region, whether gaged or ungaged, by use of the equation:

$$Q_T=M+K_{gT}S, \tag{9}$$

where

Q_T is the discharge corresponding to a recurrence interval of T years,
M is the mean of the peak discharges,
S is the standard deviation of the peak discharges, and
K_{gT} is a characteristic of the Pearson type III distribution; for the purpose of this report it may be defined as a coefficient corresponding to coefficient of skew equal to g and a recurrence interval of T years. (Table 1 gives values of K corresponding to selected values of g and recurrence interval.)

GAMMA DISTRIBUTION

The gamma distribution is a special case of the Pearson type III distribution where the locus parameter is zero (Thom, 1958, p. 117). It is also similar to the chi-square distribution that is commonly used in statistical tests. The gamma distribution has two parameters—the arithmetic mean, M, and a shape parameter, C. The parameter, C, is a function of log (M/Mg), where Mg is the geometric mean of the array (fig. 2). The use of the gamma distribution in hydrologic

TABLE 1.—*Pearson type III recurrence curve data,* K_{gT} *values, in standard deviations from the mean*

Skew coefficients (g)	Recurrence interval, in years						
	1.1	2	5	10	20	50	100
3.82 [1]						[2] 3.25	[2] 4.28
3.0	−0.65	−0.40	0.45	1.18	2.02	3.13	4.02
2.8	−.69	−.38	.48	1.20	2.02	3.09	3.95
2.6	−.74	−.37	.51	1.23	2.01	3.05	3.87
2.4	−.80	−.35	.55	1.25	2.01	3.00	3.78
2.2	−.86	−.33	.58	1.28	2.01	2.96	3.70
2.0	−.91	−.31	.61	1.30	2.00	2.91	3.60
1.9	−.93	−.30	.62	1.31	1.99	2.88	3.55
1.8	−.96	−.28	.64	1.32	1.98	2.85	3.50
1.7	−.99	−.27	.66	1.32	1.97	2.81	3.45
1.6	−1.02	−.25	.68	1.33	1.96	2.77	3.40
1.5	−1.04	−.23	.70	1.33	1.94	2.73	3.34
1.4	−1.07	−.22	.72	1.34	1.93	2.70	3.28
1.3	−1.09	−.21	.73	1.35	1.92	2.67	3.22
1.2	−1.11	−.19	.74	1.35	1.90	2.63	3.15
1.1	−1.13	−.18	.75	1.34	1.89	2.58	3.09
1.0	−1.16	−.16	.76	1.34	1.88	2.54	3.03
0.9	−1.18	−.15	.77	1.34	1.86	2.50	2.96
0.8	−1.20	−.13	.78	1.34	1.84	2.45	2.90
0.7	−1.22	−.12	.79	1.33	1.82	2.41	2.84
0.6	−1.24	−.10	.80	1.33	1.80	2.36	2.77
0.5	−1.26	−.08	.81	1.32	1.77	2.31	2.70
0.4	−1.27	−.07	.82	1.32	1.75	2.26	2.62
0.3	−1.29	−.05	.82	1.31	1.73	2.21	2.55
0.2	−1.31	−.03	.83	1.30	1.70	2.16	2.48
0.1	−1.32	−.02	.84	1.29	1.67	2.11	2.40
0	−1.33	0	.84	1.28	1.64	2.05	2.33

[1] Regional value computed for San Diego area.
[2] Extrapolated for the purpose of this report.

probability studies has been discussed by Alexander (1962); its application in a flood-frequency study is described below. Only those stations with 10 or more years of peak-discharge record are used in the analysis.

1. Peak-discharge data within the base period are tabulated for each station in a region of probable hydrologic homogeneity.
2. The base-period arithmetic mean (M) for each station in the array is computed, as explained in the earlier discussion of the Gumbel method.
3. The base-period geometric mean (Mg) for each station in the array is computed, as explained in the earlier discussion of the logarithmic normal distribution. (Mg is the antilog of the mean of the logarithms of the peak discharges.)
4. Log (M/Mg) is first computed for each station, and is then used with the curve given in figure 2 to obtain values of the shape parameter, C.
5. The base-period arithmetic mean (M) and the shape parameter (C) are regionalized by correlation with basin and climatologic parameters in the homogeneous region.
6. The regionalized values of M and C are then used with a table of chi-square to obtain the flood-frequency curve for any site in the region whether gaged or ungaged. A table of chi-square can be

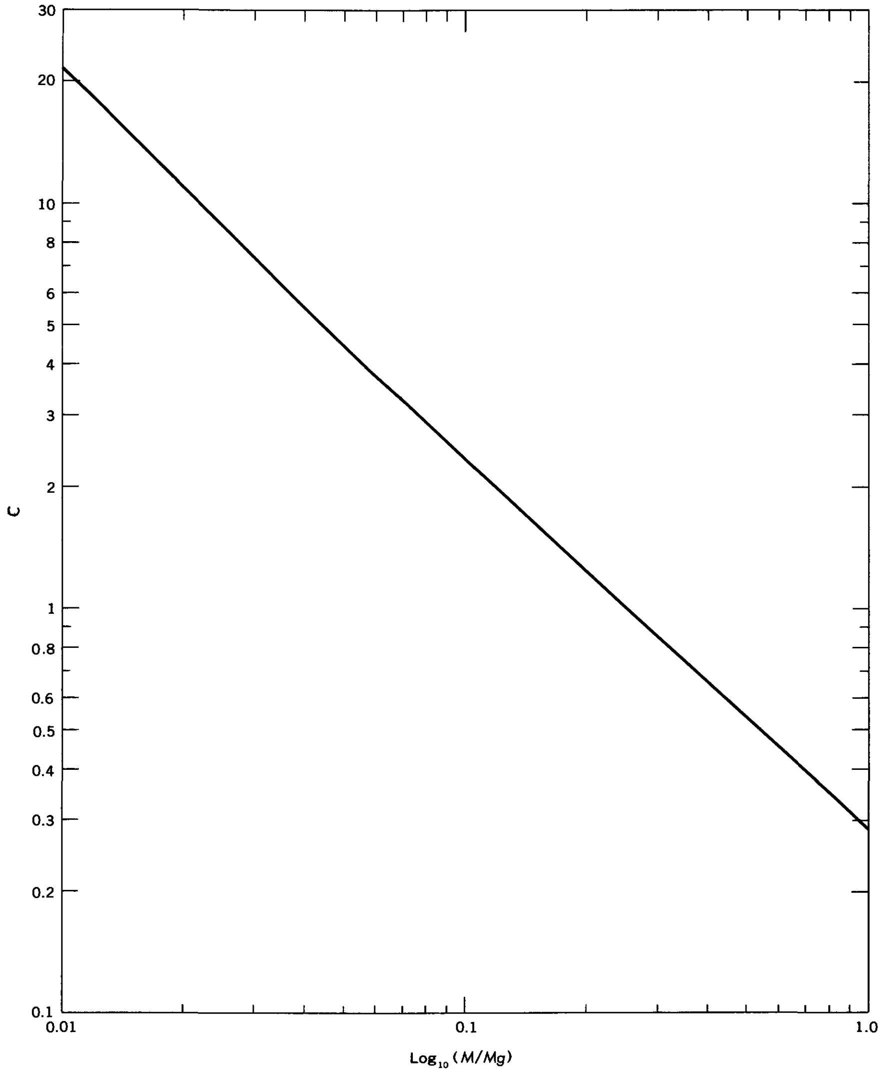


FIGURE 2.—Relation of C in the gamma distribution to log (M/Mg).

found in any handbook of statistical tables (for example, Arkin and Colton, 1950, p. 121). The number of degrees of freedom (n) to be used in the table is equal to $2C$. Each value of chi-square corresponding to $2C$ degrees of freedom is multiplied by $M/2C$ to give the ordinates of the flood-frequency curve.

7. The use of both logarithms and natural values of discharge in this method is confusing, at first glance. Logarithms are used only in the computation of C; natural values of the peak discharges are used in all other computations.

ANALYSIS OF FLOOD FREQUENCY IN THE SAN DIEGO AREA

A study of flood-frequency relations for coastal streams in the San Diego area in southwestern California was recently made by the California Department of Water Resources (1963). According to Fenneman (1931) the entire region drained by these streams lies in a single physiographic province, the Lower California province. Except for a narrow coastal plain that is 10–15 miles wide, the region is mostly mountainous having a maximum altitude of about 6,500 feet at the eastern divide. The precipitation pattern is distinctly seasonal, and about 75 percent of the rainfall occurs during the 4 months December through March. Average annual precipitation ranges from 8 inches along the coast to 45 inches at the highest altitudes. The region is generally subhumid, however, and of the 18 basins investigated, 13 had an average annual precipitation of between 15 and 21 inches. Drainage areas at the study sites range from 24 to 740 square miles.

The locations of the 18 gaging stations used in the study are shown in plate 1. The numbers identfying these stations on the map and in the tables and figures that follow are those used in the State flood-frequency report. Figure 3 shows the period of annual peak-discharge record at each of the stations, and table 2 summarizes the basin and climatologic parameters that were considered in the State study. In the published study the flood magnitude-frequency relations for the streams were analyzed by both the index-flood method and the multiple-correlation method. These two analyses are briefly summarized in the pages that follow, and in addition, analyses are made using the four statistical distributions discussed earlier—logarithmic normal, extreme-value probability, Pearson type III, and gamma. In all analyses the entire study area was considered a hydrologically homogeneous unit.

INDEX-FLOOD METHOD

In analyzing the annual flood data by the index-flood method, the State followed standard U.S. Geological Survey procedures. The base period selected for use was the 55-year period 1906–60. Annual peak discharges needed to complete the peak-discharge array for the base period for each of the 18 gaging stations were obtained by graphical correlation between stations. The discharges at each station were then ranked in order of magnitude, and the recurrence interval for each observed peak was computed by applying equation 1. Throughout most of the San Diego area the flood of 1916 was the greatest during the base period. A study of historic records indicated the peak

TABLE 2.—*Summary of hydrologic parameters for basins in the San Diego area*

Index No.	Gaging station	Drainage area (sq mi) (*A*)	Drainage basin shape factor (*Sh*)	Channel slope (ft per ft) (*S*)	Rainfall intensity (inches per day) (*I*)	Mean annual basinwide precipitation (inches) (*P*)	Mean annual basin loss (inches) (*L*)	Channel storage factor (*St*)
1	Santa Margarita River at Ysidora	740	0. 63	0. 0097	7. 50	16. 57	15. 63	7. 020
2	Santa Ysabel Creek at Sutherland Dam	57	. 63	. 0301	8. 05	28. 84	24. 23	1. 415
3	Temecula Creek at Vail Dam	319	1. 18	. 0202	9. 23	16. 24	15. 55	2. 850
4	Murrieta Creek at Temecula	220	1. 21	. 0066	6. 68	15. 16	14. 42	13. 380
5	San Juan Creek near San Juan Capistrano	110	. 67	. 0243	6. 38	19. 08	17. 60	3. 945
7	Santa Maria River near Ramona	57	. 77	. 0090	5. 45	19. 38	17. 67	7. 630
8	Sweetwater River near Descanso	44	. 69	. 0160	10. 22	28. 39	23. 80	2. 350
9	Arroyo Trabuco near San Juan Capistrano	36	. 37	. 0226	6. 30	19. 74	18. 02	3. 550
12	Santa Margarita River near Temecula	592	1. 03	. 0142	7. 93	16. 37	15. 63	5. 650
13	Santa Margarita River near Fallbrook	645	. 81	. 0132	7. 58	16. 65	15. 84	3. 870
14	Guejito Creek near San Pasqual	24	. 59	. 0229	6. 35	20. 54	18. 20	2. 500
15	San Onofre Creek near San Onofre	35	. 83	. 0526	6. 00	14. 76	14. 20	1. 560
16	Santa Ysabel Creek near Ramona	110	. 64	. 0313	6. 65	26. 68	22. 67	1. 410
17	Jamul Creek near Jamul	72	1. 13	. 0291	4. 75	17. 43	16. 35	1. 290
18	San Luis Rey River at Oceanside	348	. 52	. 0100	5. 93	20. 63	18. 89	12. 720
19	San Luis Rey River near Bonsall	306	. 65	. 0130	6. 00	21. 75	19. 98	8. 290
20	San Luis Rey River at Monserate Narrows	174	. 60	. 0171	6. 63	25. 38	22. 00	7. 510
21	San Diego River near Santee	380	. 64	. 0093	7. 35	20. 39	17. 81	8. 980

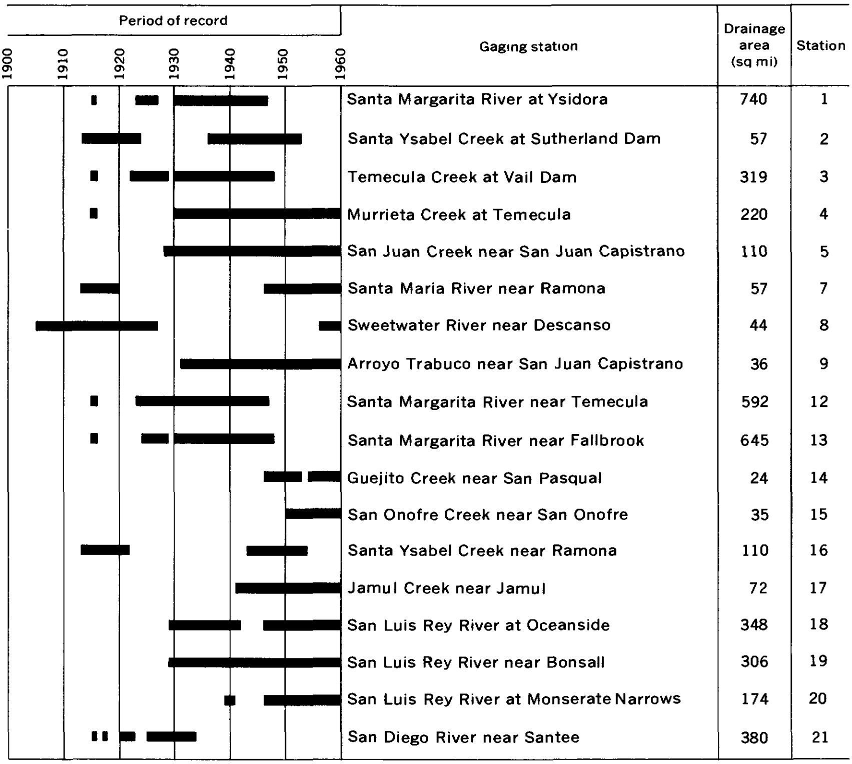


FIGURE 3.—Period of peak-discharge record for gaging stations in the San Diego area.

discharges of this flood to be of approximately the same magnitude as those of the flood of 1862, the greatest previously known. Therefore, a peak discharge equivalent to that of 1916 was assumed to have an order of magnitude of 1 and 2 in about 100 years, giving it recurrence intervals of both 100 years and 50 years. A flood-frequency curve was then drawn for each station by first plotting annual peak discharge in relation to recurrence interval and then fitting a smooth curve by eye to the plotted points. Logarithmic extreme-value probability graph paper was used for this purpose. The upper end of the curve generally fell between the two plotted first-order points. Figure 4 shows an example of a station flood-frequency curve developed by the procedure just outlined.

The discharges corresponding to recurrence intervals of 2.33 years ($Q_{2.33}$) and 10 years (Q_{10}) were selected from each of the station flood-frequency curves for use in a homogeneity test. This test indicated that all 18 stations could be used in constructing a dimensionless flood-frequency curve for the entire region. To construct this curve it was first necessary to divide the observed peak discharges at each station

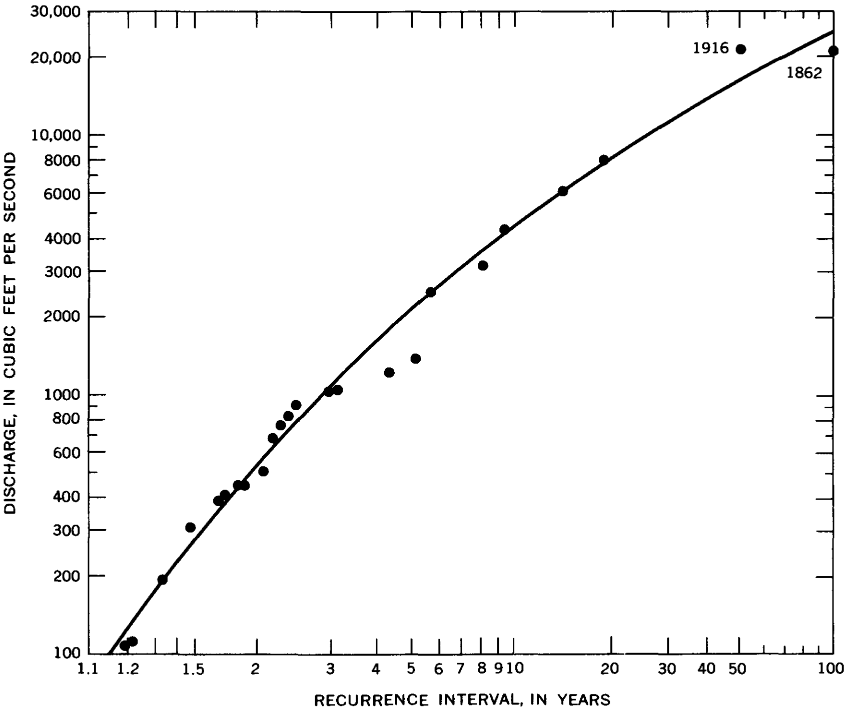


FIGURE 4.—Flood-frequency curve for Santa Ysabel Creek at Sutherland Dam (sta. 2).

by $Q_{2.33}$ so that each discharge was expressed as a dimensionless ratio. The median of the 18 peak-discharge ratios for each recurrence interval was then plotted on extreme-value probability graph paper and a smooth curve was fitted by eye to the plotted points (fig. 5). In a final step, the values of $Q_{2.33}$, picked from the individual station flood-frequency curves, were correlated with the parameters listed in table 2. Only drainage area was found to be significant; the correlation is shown in figure 6.

To obtain the flood-frequency curve for an ungaged site in the study region, figure 6 is first entered with the drainage area for the ungaged site. The corresponding mean annual flood ($Q_{2.33}$), or index flood, is then read from the curve. The value of $Q_{2.33}$ is used with the frequency curve in figure 5 to compute the ordinates of the flood-frequency curve for the ungaged area.

Values of the 50-year flood (Q_{50}) and the 100-year flood (Q_{100}) were computed from curves given in figures 5 and 6 for drainage areas corresponding in size to those of the 18 stations in the region. These values have been entered in column 4 of tables 3 and 4, respectively, for comparison with the results obtained by other methods. In column 3 of these tables are listed the values of Q_{50} and Q_{100}, picked from the individual station curves, such as the curve shown in figure 4.

MULTIPLE CORRELATION

The same 18 stations in the San Diego area were also analyzed by the California Department of Water Resources (1963), using the multiple-correlation method. The 55-year base period, 1906–60, was again used, and the entire region was considered hydrologically homogeneous. From the individual station flood-frequency curves (such as the curve given in fig. 4) constructed for the index-flood method of analysis, discharges corresponding to various recurrence intervals were read. For this study we are interested only in the discharges corresponding to recurrence intervals of 50 years (Q_{50}) and 100 years (Q_{100}). All values of Q_{50} were correlated with the basin and climatologic parameters shown in table 2; similar correlations were made for values of Q_{100}. It was found that the only statistically

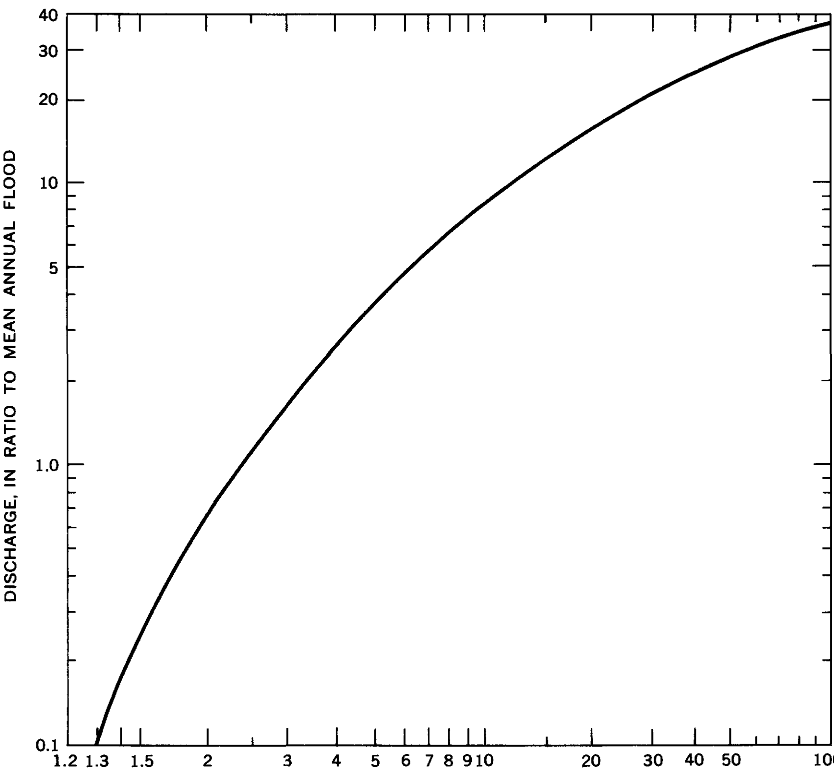


FIGURE 5.—Dimensionless regional flood-frequency curve for the San Diego area (index-flood method).

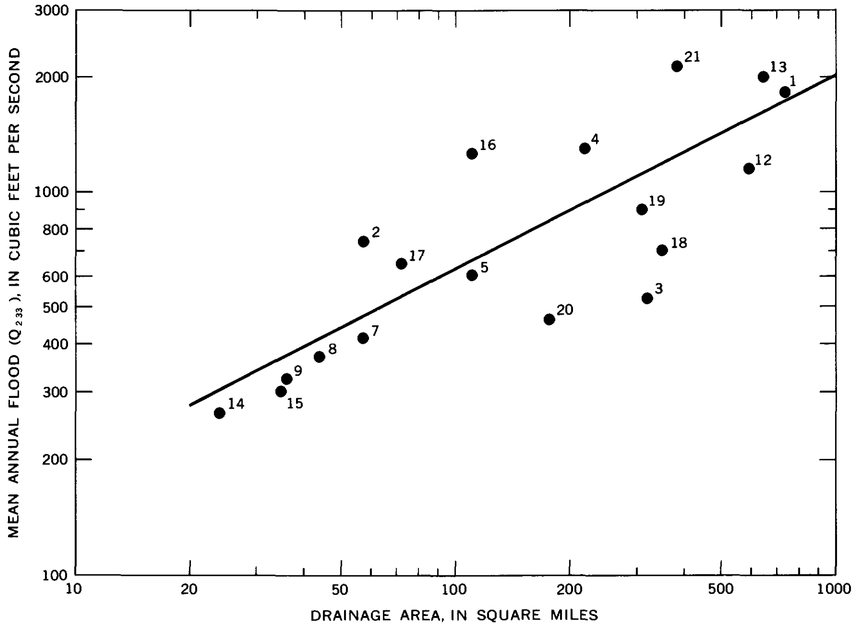


FIGURE 6.—Relation of mean annual flood to drainage area in the San Diego area (index-flood method).

significant parameters were drainage area (A) and a drainage basin shape factor (Sh). This shape factor is defined as the ratio of the diameter of a circle of area equal to basin area to the length of the basin measured parallel to the principal channel. The following regression equations were obtained:

$$Q_{50}=1016A^{0.59}Sh^{-0.44} \qquad (10)$$

and

$$Q_{100}=1288A^{0.60}Sh^{-0.57}. \qquad (11)$$

For each equation the coefficient of correlation was 0.954.

For later comparison in the section entitled, "Discussion of results of the analysis," the values of Q_{50} and Q_{100}, computed for the 18 stations by application of these equations, have been entered in column 6 of tables 3 and 4, respectively.

LOGARITHMIC NORMAL DISTRIBUTION

A regional flood-frequency curve for the San Diego region was computed by fitting a logarithmic normal distribution to the original base data. This was done to compare the results with those obtained by other methods. The first step in the computation was to convert the natural values of peak discharge to logarithms. None of the 18 gaging stations had peak-discharge records that were complete for the

TABLE 3.—Q_{50} *for streams in the San Diego area as determined from graphically derived flood-frequency curves and by various methods of regional flood-frequency analysis*

Index No.	Station	Q_{50} from graphically-derived station curves (cfs)	Index-flood method		Multiple correlation		Logarithmic normal distribution		Extreme-value probability distribution		Pearson type III distribution		Gamma distribution	
			Q_{50} (cfs)	Percent difference from col. 3	Q_{50} (cfs)	Percent difference from col. 3	Q_{50} (cfs)	Percent difference from col. 3	Q_{50} (cfs)	Percent difference from col. 3	Q_{50} (cfs)	Percent difference from col. 3	Q_{50} (cfs)	Percent difference from col. 3
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Santa Margarita River at Ysidora	54,000	50,800	−6	60,000	+11	99,800	+85	28,700	−47	34,600	−36	28,900	−46
2	Santa Ysabel Creek at Sutherland Dam	17,000	13,300	−22	13,500	−21	15,500	−9	8,530	−50	10,400	−39	7,590	−55
3	Temecula Creek at Vail Dam	24,000	32,900	+37	28,000	+17	53,200	+122	19,300	−20	23,400	−2	18,600	−22
4	Murrieta Creek at Temecula	24,500	27,300	+11	22,000	−10	42,300	+73	16,200	−34	19,700	−20	15,200	−38
5	San Juan Creek near San Juan Capistrano	26,000	18,900	−27	19,000	−27	25,400	−2	11,600	−55	14,100	−46	10,600	−59
7	Santa Maria Creek near Ramona	9,300	13,300	+43	12,500	+34	15,500	+67	8,530	−8	10,400	+12	7,590	−18
8	Sweetwater River near Descanso	10,500	11,600	+10	11,000	+5	13,000	+24	7,530	−28	9,160	−13	6,600	−37
9	Arroyo Trabuco near San Juan Capistrano	13,400	10,500	−22	13,000	−3	11,400	−15	6,830	−49	8,300	−38	5,960	−56
12	Santa Margarita River near Temecula	36,500	45,400	+24	42,500	+16	83,600	+129	25,700	−30	31,100	−15	25,700	−30
13	Santa Margarita River near Fallbrook	50,000	47,700	−5	50,000	0	89,000	+78	26,900	−46	32,600	−35	26,900	−46
14	Guejito Creek near San Pasqual	5,500	8,520	+55	8,400	+53	8,590	+56	5,670	+3	6,900	+25	4,790	−13
15	San Onofre Creek near San Onofre	9,900	10,400	+5	8,900	−10	11,100	+12	6,730	−32	8,180	−17	5,840	−41
16	Santa Ysabel Creek near Ramona	23,600	18,900	−20	19,500	−17	25,400	+8	11,600	−51	14,100	−40	10,600	−55
17	Jamul Creek near Jamul	13,000	15,200	+17	12,000	−8	18,600	+43	9,510	−27	11,600	−11	8,580	−34
18	San Luis Rey River at Oceanside	32,000	34,400	+8	42,000	+31	56,200	+76	19,900	−38	24,100	−25	19,400	−39
19	San Luis Rey River near Bonsall	36,000	32,100	−11	36,000	0	52,400	+46	18,800	−48	22,800	−37	18,200	−49
20	San Luis Rey River at Monserate Narrows	27,000	24,000	−11	27,000	0	35,400	+31	14,500	−46	17,600	−35	13,600	−50
21	San Diego River near Santee	54,000	36,400	−33	40,000	−26	60,400	+12	21,000	−61	25,400	−53	20,400	−62

TABLE 4.—Q_{100} *for streams in the San Diego area as determined from graphically derived flood-frequency curves and by various methods of regional flood-frequency analysis*

Index No.	Station	Q_{100} from graphically-derived station curves (cfs)	Index-flood method		Multiple correlation		Logarithmic normal distribution		Extreme-value probability distribution		Pearson type III distribution		Gamma distribution	
			Q_{100} (cfs)	Percent difference from col. 3	Q_{100} (cfs)	Percent difference from col. 3	Q_{100} (cfs)	Percent difference from col. 3	Q_{100} (cfs)	Percent difference from col. 3	Q_{100} (cfs)	Percent difference from col. 3	Q_{100} (cfs)	Percent difference from col. 3
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Santa Margarita River at Ysidora	72,000	70,500	−2	86,000	+19	190,000	+164	33,600	−53	44,200	−39	35,300	−51
2	Santa Ysabel Creek at Sutherland Dam	26,000	18,500	−29	18,500	−29	27,000	+4	10,000	−62	13,200	−49	9,260	−64
3	Temecula Creek at Vail Dam	34,000	45,700	+34	37,000	+9	98,200	+189	22,600	−34	29,700	−13	22,700	−33
4	Murrieta Creek at Temecula	27,500	37,800	+37	28,500	+4	77,400	+181	19,100	−31	25,100	−9	18,600	−32
5	San Juan Creek near San Juan Capistrano	36,500	26,200	−28	26,500	−27	45,400	+24	13,700	−62	18,000	−51	12,900	−65
7	Santa Maria River near Ramona	12,800	18,500	+45	16,500	+29	27,000	+111	10,000	−22	13,200	+3	9,260	−28
8	Sweetwater River near Descanso	15,500	16,200	+5	15,500	0	22,400	+45	8,880	−43	11,700	−25	8,050	−48
9	Arroyo Trabuco near San Juan Capistrano	18,500	14,600	−21	19,000	+3	19,600	+6	8,050	−56	10,600	−43	7,270	−61
12	Santa Margarita River near Temecula	49,000	63,000	+29	57,000	+16	158,000	+222	30,100	−39	39,500	−19	31,300	−36
13	Santa Margarita River near Fallbrook	67,000	66,200	−1	68,500	+2	168,000	+151	31,600	−53	41,400	−38	32,800	−51
14	Guejito Creek near San Pasqual	7,200	11,800	+64	11,500	+60	14,600	+103	6,690	−7	8,820	+22	5,840	−19
15	San Onofre Creek near San Onofre	13,300	14,400	+8	12,000	−10	19,100	+44	7,930	−40	10,500	−21	7,120	−46
16	Santa Ysabel Creek near Ramona	30,500	26,200	−14	27,000	−11	45,400	+49	13,700	−55	18,000	−41	12,900	−58
17	Jamul Creek near Jamul	17,000	21,100	+24	15,500	−9	32,700	+92	11,200	−34	14,800	−13	10,500	−38
18	San Luis Rey River at Oceanside	52,000	47,700	−8	60,000	+15	104,000	+100	23,400	−55	30,700	−41	23,700	−54
19	San Luis Rey River near Bonsall	50,000	44,500	−11	50,000	0	96,600	+93	22,100	−56	29,000	−42	22,200	−56
20	San Luis Rey River at Monserate Narrows	39,000	33,300	−15	37,000	−5	64,200	+65	17,000	−56	22,400	−43	16,500	−58
21	San Diego River near Santee	75,000	50,400	−33	56,000	−25	112,000	+49	24,700	−67	32,400	−57	24,900	−67

entire 55-year base period, 1906–60. The next step was to select a few stations that were strategically located for use as base stations, and then to estimate the logarithms of the annual peak discharges needed to complete the 55-year array for these stations. Three of the stations having comparatively long records, stations 2, 3, and 8, were chosen as base stations, and the required individual logarithms of peak discharge were computed by use of equation 7. (Note that logarithms are used at this time in equation 7 and not natural values.) The use of equation 7, instead of a regression equation, results in a less biased estimate of the standard deviation of the 55-year array of annual peak discharges, not all of which are observed.

The mean and the standard deviation for the 55-year array at each of the three base stations were then computed. After this was done, the mean and the standard deviation of each of the remaining 15 short-term records were first computed and then adjusted to the 55-year base period by correlation with a base station. Equations 3 and 4 were then applied to the results. At this stage of the computations the base-period mean (M) and standard deviation (S) were available for all 18 stations. The values of M and S, still in logarithmic units, were next regionalized by correlation with the basin and climatologic parameters in table 2. Of these parameters, only drainage area was statistically significant. The two graphs in figure 7 show the relations between (1) drainage area and M, the mean of the logarithms of peak discharge; and (2) drainage area and S, the standard deviation of the logarithms of discharge. By use of these two graphs and equation 5, a flood-frequency curve can be computed for any site in the region. Specific equations for computing Q_{50} and Q_{100}, in logarithmic units, are:

$$Q_{50} = \mathrm{M} + 2.05S \tag{12}$$

and

$$Q_{100} = M + 2.33S. \tag{13}$$

Values of Q_{50} and Q_{100}, in logarithms, were computed from curves given in figure 7 and by application of equations 12 and 13, respectively, for drainage areas corresponding in size to those of the 18 gaging stations. The antilogarithms of Q_{50} and Q_{100} have been entered in column 8 of tables 3 and 4, respectively, for later comparison.

EXTREME-VALUE PROBABILITY DISTRIBUTION OR GUMBEL METHOD

A regional flood-frequency study of the San Diego region was also made by fitting an extreme-value probability distribution to the original base data. In this analysis natural values of the discharges were used, and not their logarithms. Stations 2, 3, and 8 were again

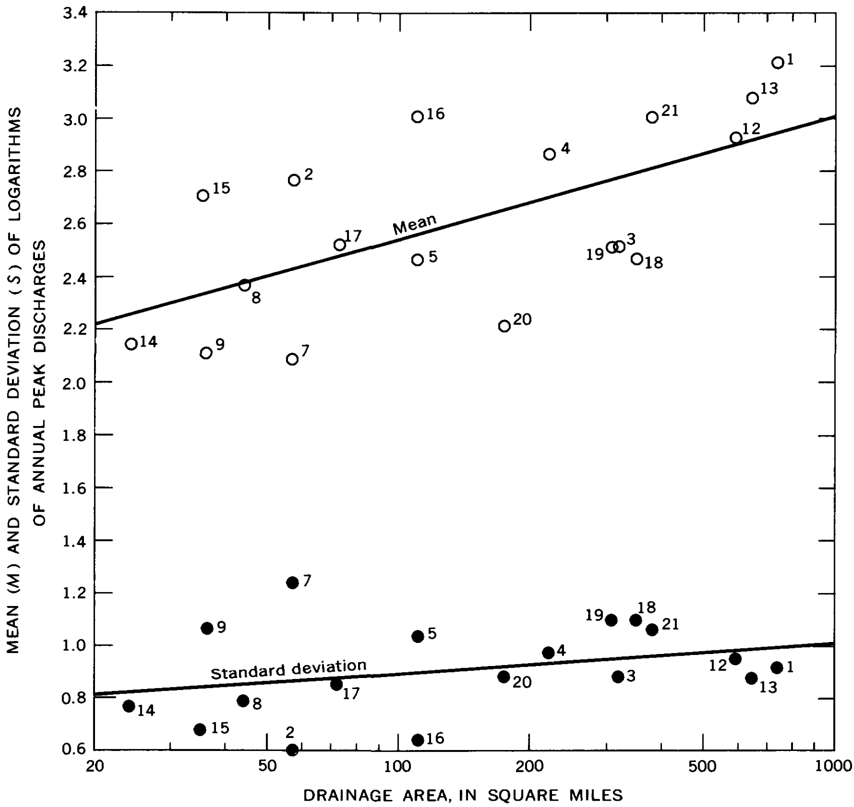


Figure 7.—Relation of mean and standard deviation of logarithms of annual peak discharges to drainage area in the San Diego area.

made the base stations for the study and their records were completed for the 55-year base period by use of equation 7. The mean and the standard deviations for the 3 base stations and the 15 short-term stations were then computed. These statistics for the short-term records were adjusted to the 55-year base period by correlations with a base station. Equations 3 and 4 were then applied to the results. (Note that natural values are used in this method in equations 3 and 4.) The resulting base-period values of the mean (M) and standard deviation (S) were regionalized by correlation with the basin and climatologic parameters listed in table 2. Again, drainage area was the only statistically significant parameter. The two graphs in figure 8 show the relation between (1) drainage area and M, and (2) drainage area and S. By use of these two graphs and equation 6, a flood-frequency curve can be computed for any site in the region. Specific equations for computing Q_{50} and Q_{100} are:

$$Q_{50}=M+2.59S \tag{14}$$

and

$$Q_{100}=M+3.14S. \tag{15}$$

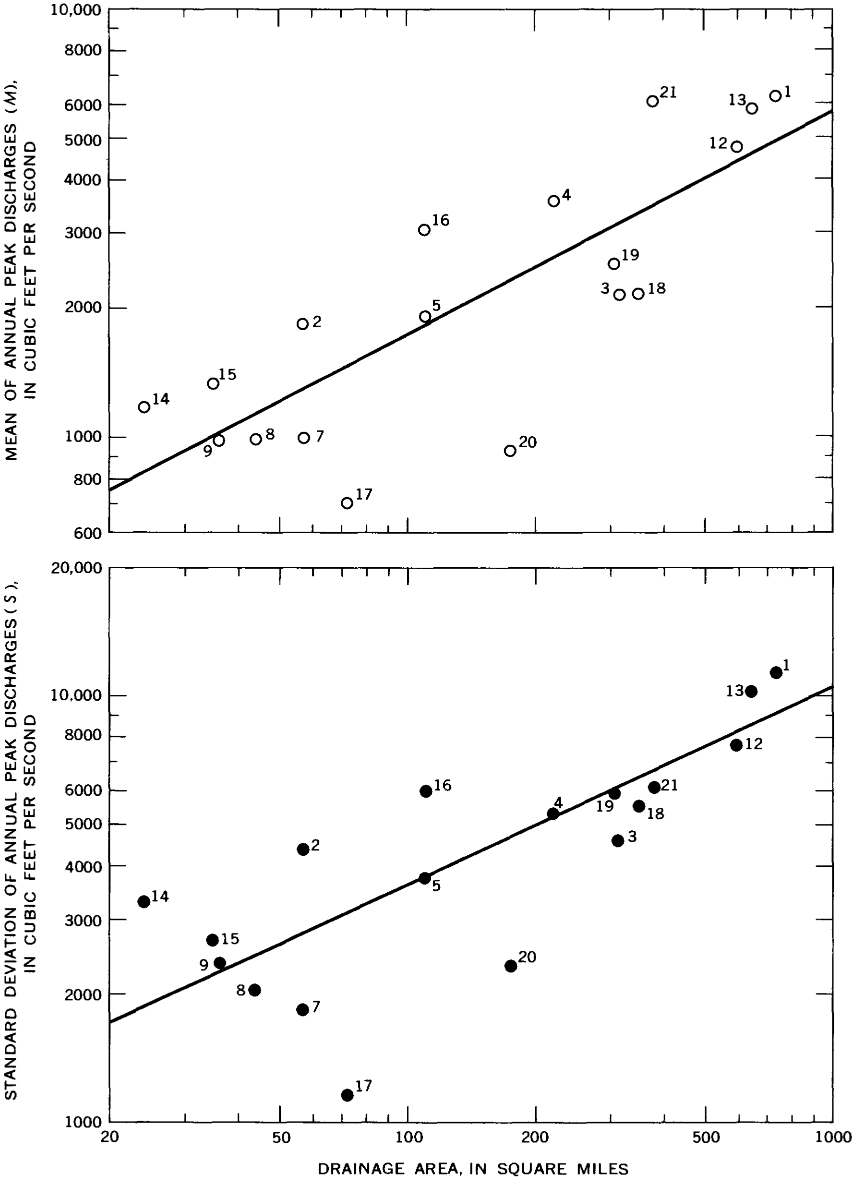


FIGURE 8.—Relation of mean and standard deviation of annual peak discharges to drainage area in the San Diego area.

770-401 O - 65 - 5

Values of Q_{50} and Q_{100} were computed from curves given in figure 8 and by application of equations 14 and 15, respectively, for drainage areas corresponding in size to those of the 18 gaging stations. These values are listed in column 10 of tables 3 and 4, respectively, for later comparison.

PEARSON TYPE III DISTRIBUTION

A fifth flood-frequency analysis was made for the San Diego area streams by fitting a Pearson type III distribution to the peak-discharge data. The computation and regionalization of the base-period mean (M) and standard deviation (S) for this study are identical with those for the extreme-value probability analysis. Consequently, the graphs in figure 8 were applicable to this analysis. The next step was to determine the coefficient of skew for each station. This was done by first using equation 7 to compute the annual peak discharges needed to complete the 55-year array at each short-term station and then using equation 8 to compute the coefficient of skew (g) for each station. The individual values of g ranged from 2.39 to 5.35 but showed no correlation with basin or climatologic parameters. Consequently, the average value of g, 3.82, was assumed to be its regional value. A flood-frequency curve can be computed for any site in the region, gaged or ungaged, by first entering figure 8 with drainage area to obtain M and S and then using equation 9 and table 1 with the regional skew of 3.82.

Values of Q_{50} and Q_{100} were computed by use of the relations and table just mentioned for drainage areas corresponding in size to those of the 18 gaging stations. These values have been entered in column 12 of tables 3 and 4, respectively, for later comparison with the results obtained by the other methods of analysis.

GAMMA DISTRIBUTION

The gamma distribution was also used to analyze the flood magnitude-frequency relation for streams in the San Diego area. A prerequisite to the use of this method was a determination of the base-period arithmetic mean (M) and geometric mean (Mg) for each of the 18 stations. Values of M were available from the previous use of the Gumbel-method analysis, and the logarithms of Mg were available from the previous use of the logarithmic normal distribution. The logarithms of Mg were then converted to natural values. The logarithm of the ratio M/Mg was next computed for each station, and the curve given in figure 2 was then used to obtain corresponding values of C. The next step was to regionalize the values of M and C. This had been done for M in the Gumbel method of analysis and the results are found in figure 8. The individual values of C ranged from

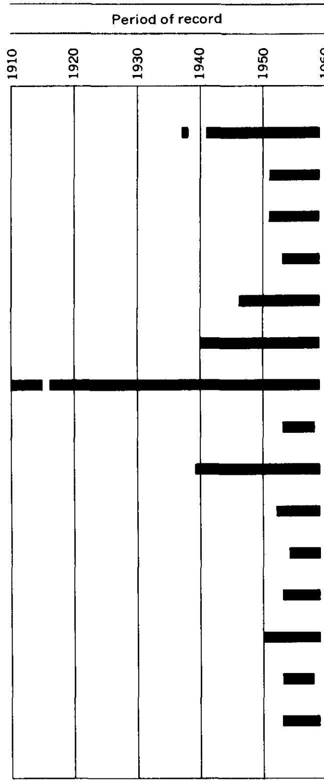
0.31 to 0.89, but showed no correlation with basin or climatologic parameters. Consequently, the average value of C, 0.43, was assumed to be its regional value. The regional values of M and C can be used with a table of chi-square to compute a flood-frequency curve for any site in the region, whether gaged or ungaged.

Values of Q_{50} and Q_{100} were computed by use of curves given in figure 8, a regional value of C of 0.42, and a chi-square table, for drainage areas corresponding in size to those of the 18 gaging stations. These values are listed in column 14 of tables 3 and 4, respectively, for later comparison with the results obtained by the other five methods of analysis.

ANALYSIS OF FLOOD FREQUENCY IN NORTH COASTAL CALIFORNIA

A study of flood-frequency relations for coastal streams in northern California was recently made by the U.S. Geological Survey (Rantz, 1964, p. 60–73). According to Fenneman (1931), the region drained by these streams lies in two physiographic sections—the northern California Coast Ranges and the Klamath Mountains. Except for a narrow coastal plain, the region is very mountainous with many peaks above an altitude of 6,000 feet. The precipitation is distinctly seasonal and about 75 percent of it occurs during the 5 months, November through March. Snow falls in moderate amounts at altitudes above 2,000 feet, but only at altitudes above 4,000 feet does snow remain on the ground for appreciably long periods. Average annual precipitation increases from east to west, ranging from a low of 10 inches in the Shasta River valley to a high of 120 inches in the upper Smith River basin. The region as a whole is very humid; and of the 27 basins investigated, only 5 had an average annual precipitation of less than 50 inches. Drainage areas at the study sites ranged from 6 to 12,000 square miles.

The locations of the 27 gaging stations used in the study are shown in plate 2. The numbers identifying these stations on the map and in the tables and figures that follow are those listed in the Geological Survey study. Figure 9 shows the period of annual peak-discharge record at each of the stations, and table 5 lists the two hydrologic parameters—drainage area and mean annual precipitation—that were found to have a significant effect on the flood magnitude-frequency relation. In the Geological Survey study this relation was analyzed by the index-flood method. That analysis is summarized in the pages that follow, and in addition, analyses were made by five other methods—multiple correlation, logarithmic normal distribution, extreme-value probability distribution, Pearson type III distribution, and the



Gaging station	Drainage area (sq mi)	Station
Coast Ranges		
East Fork Russian River near Calpella	93	4615
Middle Fork Eel River below Black Butte River near Covelo	367	4730
Eel River below Dos Rios	1481	4740
North Fork Eel River near Mina	251	4745
South Fork Eel River near Branscomb	43.9	4755
South Fork Eel River near Miranda	537	4765
Eel River at Scotia	3113	4770
Van Duzen River near Dinsmores	80.2	4775
Van Duzen River near Bridgeville	214	4785
Yager Creek near Carlotta	127	4790
Jacoby Creek near Freshwater	6.07	4800
Mad River near Forest Glen	144	4805
Mad River near Arcata	485	4810
Redwood Creek near Blue Lake	67.5	4815
Redwood Creek at Orick	278	4825

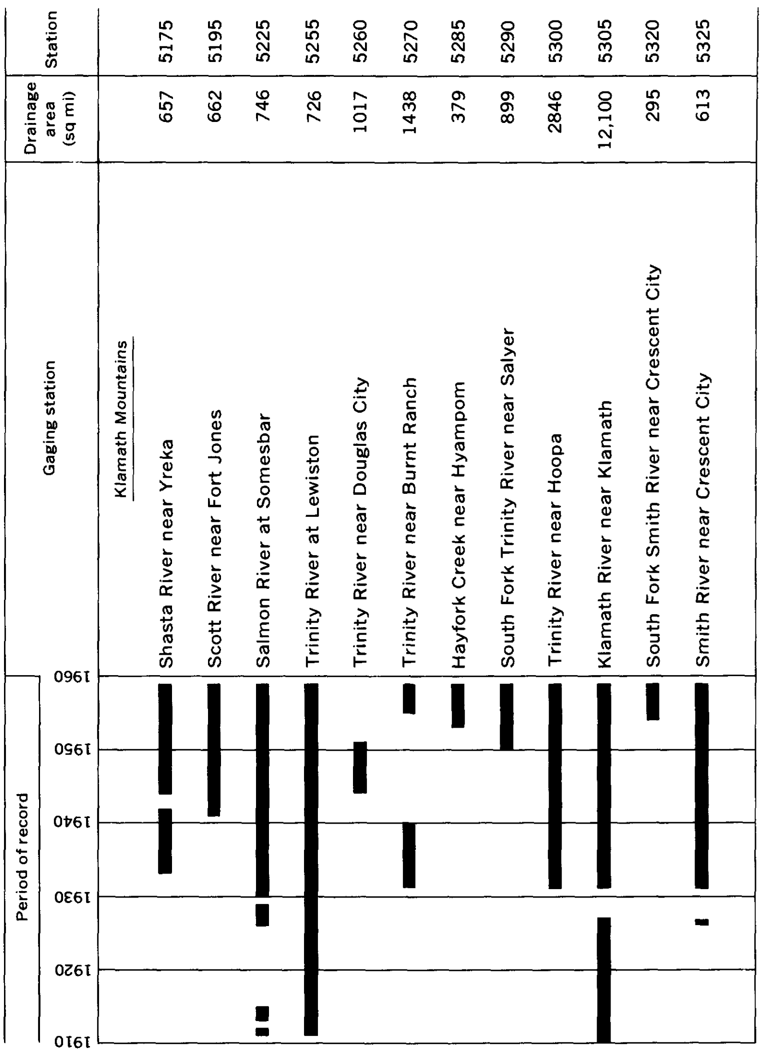


FIGURE 9.—Period of peak-discharge record for gaging-stations in north coastal California.

gamma distribution. In differentiating regional characteristics of $Q_{2.33}$ in the index-flood method and in all subsequent analyses of the data for north coastal California, each of the two physiographic sections—the Coast Ranges and the Klamath Mountains—was considered separately as a hydrologic entity.

TABLE 5.—*Summary of pertinent hydrologic parameters for basins in north coastal California*

[Asterisk indicates long-term station]

Index No.	Gaging station	Drainage area (sq mi)	Mean annual basinwide precipitation (inches)
	Coast Ranges		
*4615	East Fork Russian River near Calpella	93.0	40
4730	Middle Fork Eel River below Black Butte River, near Covelo	367	60
4740	Eel River below Dos Rios	1,481	53
4745	North Fork Eel River near Mina	251	58
*4755	South Fork Eel River near Branscomb	43.9	79
*4765	South Fork Eel River near Miranda	537	70
*4770	Eel River at Scotia	3,113	59
4775	Van Duzen River near Dinsmores	80.2	74
*4785	Van Duzen River near Bridgeville	214	72
4790	Yager Creek near Carlotta	127	60
4800	Jacoby Creek near Freshwater	6.07	54
4805	Mad River near Forest Glen	144	60
4810	Mad River near Arcata	485	64
4815	Redwood Creek near Blue Lake	67.5	80
4825	Redwood Creek at Orick	278	80
	Klamath Monntains		
*5175	Shasta River near Yreka	[1] 657	19
*5195	Scott River near Fort Jones	662	33
*5225	Salmon River at Somesbar	746	57
*5255	Trinity River at Lewiston	726	59
5260	Trinity River near Douglas City	1,017	56
*5270	Trinity River near Burnt Ranch	1,438	57
5285	Hayfork Creek near Hyampom	379	43
5290	South Fork Trinity River near Salyer	899	50
*5300	Trinity River near Hoopa	2,846	55
*5305	Klamath River near Klamath	12,100	42
5320	South Fork Smith River near Crescent City	295	116
*5325	Smith River near Crescent City	613	111

[1] Actual drainage area above gage is 796 sq mi, but 139 sq mi above Dwinnell Reservoir is noncontributing.

INDEX-FLOOD METHOD

In analyzing the annual flood data by the index-flood method, Rantz (1964) used a 28-year base period, 1932–59. Of the 27 gaging stations studied, only 13 had records with 10 or more years within the base period. Annual peak discharges needed to complete the discharge array for the base period for these 13 stations were obtained by graphical correlation. The discharges at each of the stations were then ranked in order of magnitude, and the recurrence interval for each observed peak was computed by applying equation 1. A flood-frequency curve was next drawn on extreme-value probability graph paper for each of the 13 stations by first plotting annual peak discharge against recurrence interval and then fitting a smooth curve by eye to the plotted points.

It was possible to extrapolate the flood-frequency curves beyond the base period with considerable confidence because of the availability of historical records, both qualitative and quantitative, of major floods that occurred in years prior to 1932. For Klamath River at Klamath (sta. 5305), the magnitudes of all major flood peaks in the past 106 years are known and were used in the construction of the flood-frequency curve (fig. 10). For other stations, where it was known only that the flood peaks of 1956 were roughly equivalent to those of 1862 and greater than any other since at least 1854, the magnitude of

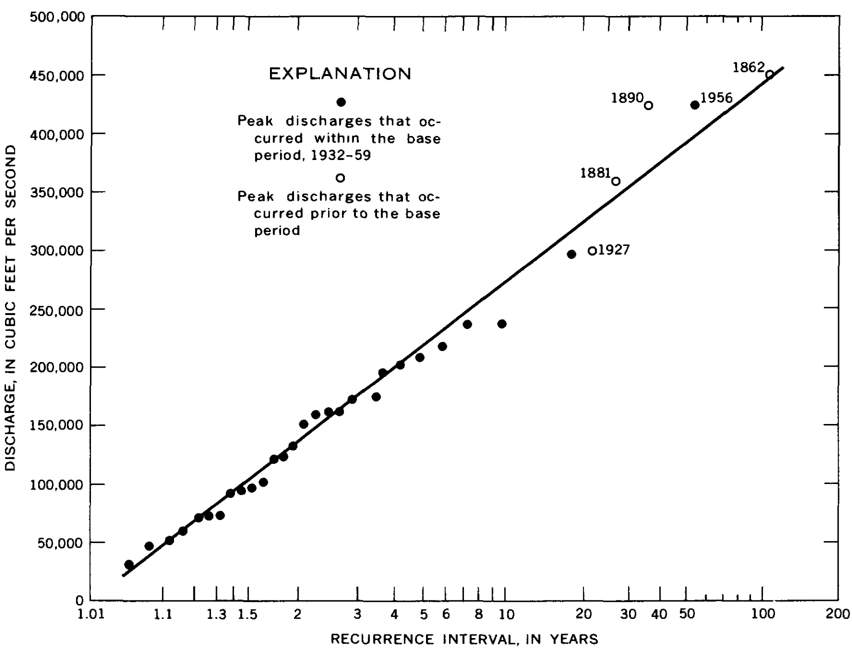


FIGURE 10.—Flood-frequency curve for Klamath River at Klamath (sta. 5305).

the 1956 peak was plotted at both 107 years and 53.5 years to indicate that this magnitude represented both the highest and second highest discharges in 106 years. The well-defined part of the flood-frequency curve for each station, generally a straight line or gentle curve, was extrapolated to the 100-year recurrence interval, with the provision that the extrapolation pass through one of these two plotted points or pass between them. An example of this extrapolation is the flood-frequency curve for Eel River at Scotia (sta. 4770), shown in fig. 11.

The slopes of the station flood-frequency curves were next tested for homogeneity. The results of this test are better understood in the light of the explanatory remarks that follow. The slope of the flood-frequency curve for northern California streams is influenced primarily by the difference in severity between the storms that cause the milder floods, such as the mean annual flood, and the storms that cause the infrequent major floods. The greater the disparity between these two types of storm, the greater the ratio of major flood peak to the mean annual flood peak and, therefore, the steeper the slope of the flood-frequency curve. Furthermore, it is almost axiomatic that the more humid the area, the less variability there is in the precipitation. Consequently, the areas closest to the coast, since they in general have the greatest precipitation, would be expected to have flood-frequency curves that show the flattest slope. Infiltration capacity has a relatively small effect on the peak discharge during major floods because these floods are generally associated with rains that last for many days, and consequently, the ground becomes well saturated and the infiltrating rain amounts to only a small percentage of the storm precipitation. Altitude may also be a factor because during these prolonged major storms there is generally some snowmelt which augments the runoff directly attributable to rainfall. Thus, the flood-frequency curves

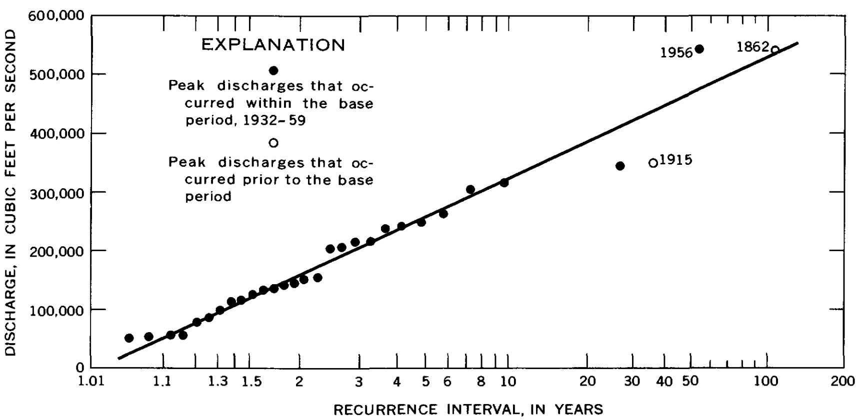


FIGURE 11.—Flood-frequency curve for Eel River at Scotia (sta. 4770).

for the basins of higher altitude in northwestern California tend to have steeper slopes.

The statistical tests for homogeneity of slope of the flood-frequency curves bore out these premises. These tests resulted in the establishment of the three areas of homogeneity shown in plate 2. The dimensions flood-frequency curves for the three subregions are plotted in figure 12. Subregion 1 has the flattest station flood-frequency curves; lying closest to the ocean, it is the most humid area and has the lowest altitude. Its dimensionless flood-frequency curve is based on the eight gaging stations in subregion 1 that had been in operation for at least 10 years. The slope of the dimensionless flood-frequency curve for subregion 2 is steeper than that of the flood-frequency curve for subregion 1 owing to the generally more variable storm precipitation and higher altitudes found in subregion 2. Only three stations have 10 or more years of record in subregion 2, and consequently the flood-frequency curve representative of this subregion lacks the high degree of confirmation obtainable from a large number of gaging stations. The fact that the dimensionless flood-frequency curve for subregion 3 has the steepest slope of the three regional curves reflects that subregion 3 is the least humid of the three subregions. Only two stations in the subregion have the requisite 10 or more years of record.

The magnitude of the mean annual flood ($Q_{2.33}$) was next investigated. For the 13 stations with 10 or more years of peak-discharge record, $Q_{2.33}$ was picked from the individual station flood-frequency

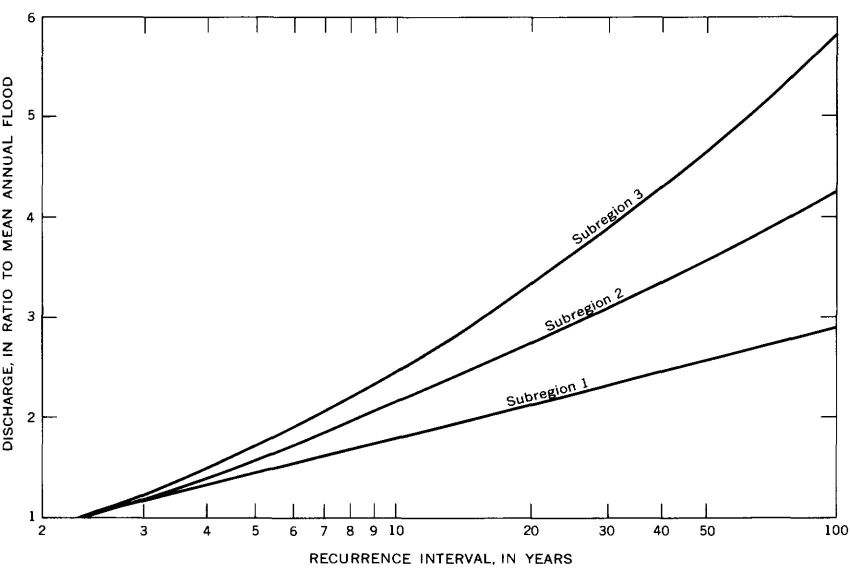


FIGURE 12.—Dimensionless regional flood-frequency curves for north coastal California (index-flood method).

curves that had been previously plotted. For the 14 stations with less than 10 years of record, correlation procedures were used to determine $Q_{2.33}$. Concurrent peak discharges for each of these short-term stations and for a nearby longer term station were correlated graphically. $Q_{2.33}$ for the short-term station was then determined from the regression line, it being the discharge corresponding to $Q_{2.33}$ for the longer-term station. The values of $Q_{2.33}$ for both long-term and short-term stations were then analyzed.

The magnitude of the mean annual flood in basins in northwestern California is related primarily to the size of drainage area and to the magnitude of the mean annual storm. Mean annual precipitation is an excellent index of the relative magnitude of the mean annual storm because the bulk of the annual precipitation in the region occurs during several general storms each year, and all stations experience the same number of general storms in any given year. Subsurface storage also exerts a significant influence on the magnitude of the mean annual flood. Surface storage, on the other hand, is a negligible factor in this study because there are no sizable lakes or reservoirs that are uncontrolled, and streams that are seriously affected by artificially regulated storage were excluded from the analysis. Because subsurface storage is related to the infiltration capacity, or the permeability of the soil and mantle rock, and because much greater permeability is associated with the Klamath Mountains than with the Coast Ranges, it is logical to expect the mean annual flood to differ in these two physiographic sections, when all other factors are equal.

In figure 13 the mean annual floods for basins in the Coast Ranges have been plotted in relation to the drainage area. Each point is labeled with (1) the number of the gaging station for identification, and (2) the mean annual precipitation for the basin upstream from the station. (Drainage area and mean annual basinwide precipitation are given in table 5.) Precipitation in the Coast Ranges basins ranges from 40 to 80 inches, and within this range no significant correlation is apparent between mean annual flood and mean annual precipitation. A straight line averaging the plotted points has the equation:

$$\text{Mean annual flood } (Q_{2.33}) = 130A^{0.91}, \tag{16}$$

where A is the drainage area in square miles.

The wide range in mean annual precipitation in the Klamath Mountains, 19–116 inches, has a very pronounced effect on the magnitude of the mean annual flood. The relation of $Q_{2.33}$ to drainage area and mean annual precipitation in this physiographic section can be expressed by a means of a family of curves, each of which has an equation similar to equation 16. The equation for this family of curves is

$$Q_{2.33} = CA^{0.91}, \tag{17}$$

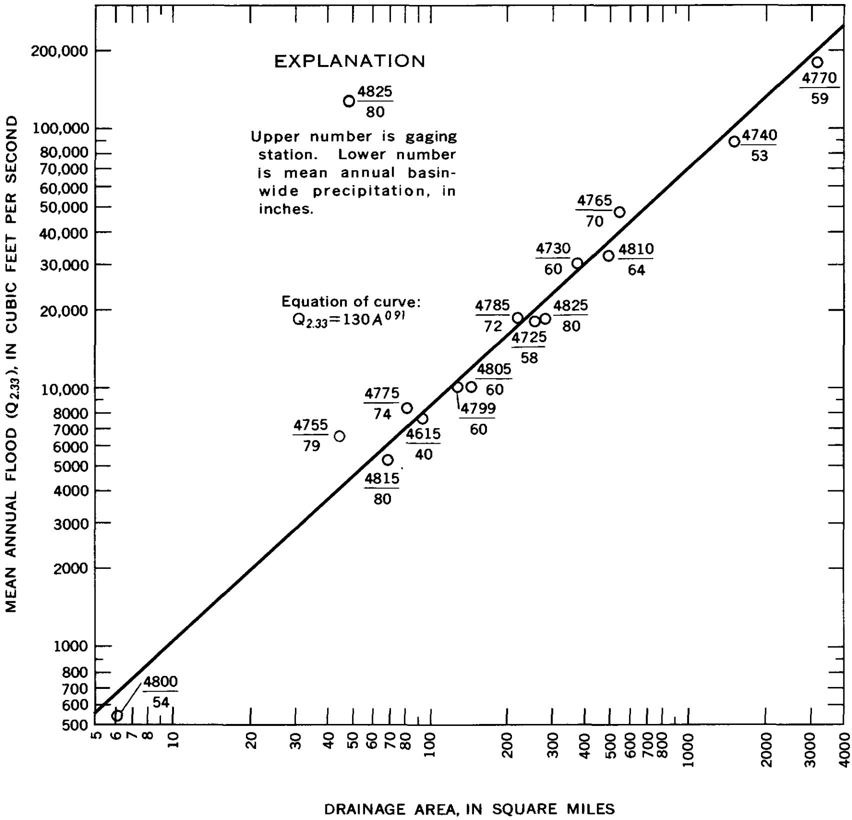


FIGURE 13.—Relation of mean annual flood to drainage area in the Coast Ranges (index-flood method).

where C is a variable that is related to mean annual precipitation. The relation of C to mean annual basinwide precipitation is shown graphically in figure 14. From a comparison of the graphs in figures 13 and 14, it is concluded that for the same size of drainage area and the same mean annual precipitation, mean annual floods are greater in the Coast Ranges than in the Klamath Mountains. The basis for this conclusion is the fact that the coefficient of 130 in the Coast Ranges formula is equivalent to C corresponding to about 90 inches of mean annual precipitation in the formula for the Klamath Mountains, yet the precipitation in the Coast Ranges ranged from only 40 to 80 inches. This result is not surprising in view of the fact that the Klamath Mountains have the more permeable soil and mantle rock.

Values of the 50-year flood (Q_{50}) and the 100-year flood (Q_{100}) were computed from curves given in figures 12–14 and equation 17 for sites whose drainage areas correspond in size and in mean annual basinwide

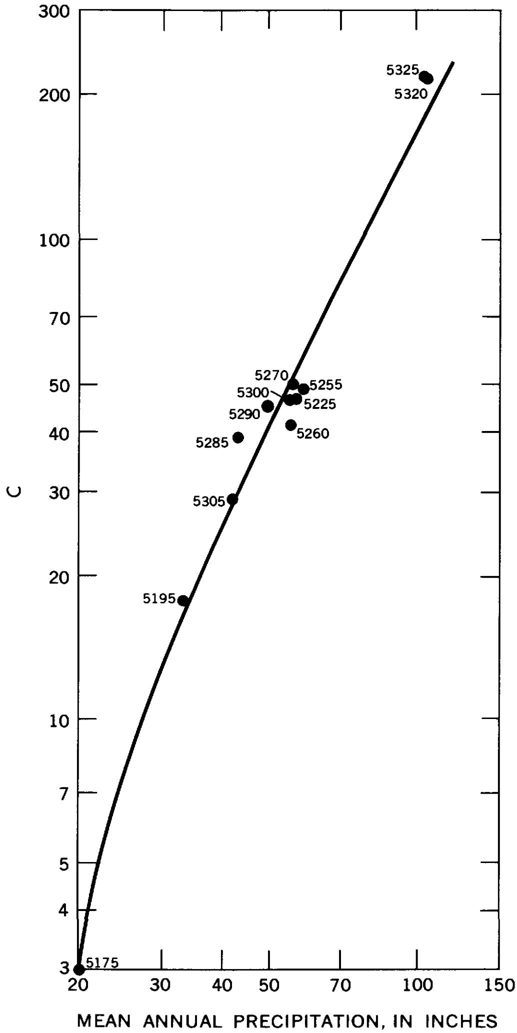


Figure 14.—Relation of *C* to mean annual basinwide precipitation in the Klamath Mountains (index-flood method).

precipitation to those of the 13 long-term stations in the region. These computed values have been entered in column 4 of tables 6 and 7, respectively, for comparison with the results obtained by other methods. Q_{50} and Q_{100} were not computed for sites similar to the 14 short-term stations because the station records were too short for use in the other methods of analysis that follow in this report. In column 3 of tables 6 and 7 are listed the values of Q_{50} and Q_{100}, picked from the individual station curves, such as the curves shown in figures 10 and 11.

TABLE 6.—Q_{50} *for streams in north coastal California as determined from graphically derived flood-frequency curves and by various methods of regional flood-frequency analysis*

Index No.	Station	Q_{50} from graphically-derived station curves (cfs)	Index-flood method		Multiple correlation		Logarithmic normal distribution		Extreme-value probability distribution		Pearson type III distribution		Gamma distribution	
			Q_{50} (cfs)	Percent difference from col. 3	Q_{50} (cfs)	Percent difference from col. 3	Q_{50} (cfs)	Percent difference from col. 3	Q_{50} (cfs)	Percent difference from col. 3	Q_{50} (cfs)	Percent difference from col. 3	Q_{50} (cfs)	Percent difference from col. 3
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	California Coast Ranges													
4615	East Fork Russian River near Calpella	19,600	20,400	+4	26,000	+33	25,200	+29	22,800	+16	23,500	+20	23,000	+17
4755	South Fork Eel River near Branscomb	18,400	10,300	−44	14,000	−24	13,600	−26	12,000	−35	12,400	−33	12,400	−33
4765	South Fork Eel River near Miranda	142,000	102,000	−28	110,000	−23	113,000	−20	102,000	−28	105,000	−26	99,500	−30
4770	Eel River at Scotia	470,000	516,000	+10	480,000	+2	481,000	+2	475,000	+1	489,000	+4	456,000	−3
4785	Van Duzen River near Bridgeville	39,300	43,900	+12	52,000	+32	50,400	+28	47,300	+20	48,700	+24	46,300	+18
	Klamath Mountains													
5175	Shasta River near Yreka	5,000	3,430	−31	6,700	+34	7,420	+48	5,700	+14	6,170	+23	5,020	0
5195	Scott River near Fort Jones	29,000	28,800	−1	20,000	−31	21,000	−28	16,800	−42	18,000	−38	12,200	−58
5225	Salmon River at Somesbar	67,000	80,500	+20	64,800	−3	65,200	−3	53,500	−20	57,600	−14	36,100	−46
5255	Trinity River at Lewiston	67,000	84,400	+26	67,800	+1	68,000	+1	55,800	−17	60,000	−10	37,300	−44
5270	Trinity River near Burnt Ranch	142,000	146,000	+3	114,000	−20	124,000	−13	96,700	−32	104,000	−27	67,000	−53
5300	Trinity River near Hoopa	170,000	183,000	+8	192,000	+13	226,000	+33	167,000	−2	180,000	+6	120,000	−29
5305	Klamath River near Klamath	395,000	382,000	−3	396,000	0	564,000	+43	367,000	−7	395,000	0	294,000	−26
5325	Smith River near Crescent City	167,000	177,000	+6	202,000	+21	178,000	+7	164,000	−2	176,000	+5	101,000	−40

TABLE 7.—Q_{100} *for streams in north coastal California as determined from graphically derived flood-frequency curves and by various methods of regional flood-frequency analysis*

Index No.	Station	Q_{100} from graphically-derived station curves (cfs)	Index-flood method		Multiple correlation		Logarithmic normal distribution		Extreme-value probability distribution		Pearson type III distribution		Gamma distribution	
			Q_{100} (cfs)	Percent difference from col. 3	Q_{100} (cfs)	Percent difference from col. 3	Q_{100} (cfs)	Percent difference from col. 3	Q_{100} (cfs)	Percent difference from col. 3	Q_{100} (cfs)	Percent difference from col. 3	Q_{100} (cfs)	Percent difference from col. 3
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	California Coast Ranges													
4615	East Fork Russian River near Calpella	22,100	23,000	+4	29,500	+33	29,400	+33	25,600	+16	26,500	+20	25,600	+16
4755	South Fork Eel River near Branscomb	20,900	11,600	−44	15,700	−25	15,800	−24	13,500	−35	13,900	−33	13,700	−34
4765	South Fork Eel River near Miranda	170,000	115,000	−32	125,000	−26	131,000	−23	115,000	−32	119,000	−30	111,000	−35
4770	Eel River at Scotia	530,000	582,000	+10	545,000	+3	560,000	+6	535,000	+1	554,000	+5	507,000	−4
4785	Van Duzen River near Bridgeville	43,600	49,500	+14	59,000	+35	58,600	+34	53,200	+22	55,000	+26	51,500	+18
	Klamath Mountains													
5175	Shasta River near Yreka	6,000	4,280	−29	8,420	+40	9,700	+62	6,650	+11	7,300	+22	5,950	−1
5195	Scott River near Fort Jones	37,500	36,000	−4	24,300	−35	26,200	−30	19,400	−48	21,200	−43	14,500	−61
5225	Salmon River at Somesbar	82,500	96,000	+16	76,400	−7	78,600	−5	61,800	−25	67,400	−18	40,500	−51
5255	Trinity River at Lewiston	76,700	101,000	+32	79,800	+4	81,800	+7	64,400	−16	70,200	−8	41,900	−45
5270	Trinity River near Burnt Ranch	170,000	174,000	+2	132,000	−22	149,000	−12	111,000	−35	122,000	−28	75,000	−56
5300	Trinity River near Hoopa	192,000	207,000	+8	220,000	+15	272,000	+42	193,000	+1	210,000	+9	134,000	−30
5305	Klamath River near Klamath	445,000	431,000	−3	445,000	0	695,000	+56	423,000	−5	461,000	+4	335,000	−25
5325	Smith River near Crescent City	186,000	200,000	+8	231,000	+24	202,000	+9	188,000	+1	204,000	+10	110,000	−41

MULTIPLE CORRELATION

In analyzing the flood magnitude-frequency relation for north coastal California by the multiple-correlation method it was necessary to eliminate 14 of the 27 gaging stations from consideration because the records for these stations were shorter than 10 years. Of the remaining 13 stations, 5 were in the Coast Ranges and 8 were in the Klamath Mountains. (See table 5.) The fact that so few records were available for study severely handicapped the analysis. Because there were only five stations in the Coast Ranges, no more than one independent variable could be used in a linear correlation, if the correlation were to have any significance at all. Because there were only eight stations in the Klamath Mountains, no more than two independent variables could be used in the correlation.

The first step in the analysis was to obtain the discharges corresponding to recurrence intervals of 50 years (Q_{50}) and 100 years (Q_{100}) from the 13 individual station flood-frequency curves. These curves (see, for example, figs. 12 and 13) had been constructed earlier for the index-flood method of analysis. The values of Q_{50} and Q_{100} were then correlated with the hydrologic parameters listed in table 5. For the Coast Ranges, drainage area (A) was by far the most significant parameter, and it was used as the sole independent variable in the correlation for the five stations in that physiographic region. Two independent variables—drainage area (A) and mean annual basin-wide precipitation (P)—were used in the multiple correlation for the eight stations in the Klamath Mountains. Regression equations for Q_{50} and Q_{100} were obtained as follows:

Coast Ranges:

$$Q_{50}=602A^{0.828} \tag{18}$$

$$Q_{100}=674A^{0.831}. \tag{19}$$

Klamath Mountains:

$$Q_{50}=0.075A^{0.866}P^{1.965} \tag{20}$$

$$Q_{100}=0.128A^{0.842}P^{1.912}. \tag{21}$$

For later comparison, the values of Q_{50} and Q_{100}, computed for the 13 stations by application of these equations, have been entered in column 6 of tables 6 and 7, respectively.

LOGARITHMIC NORMAL DISTRIBUTION

Regional flood-frequency curves for north coastal California were computed by fitting logarithmic normal distributions to the original base data for the 13 stations used in the previous multiple-correlation analysis. The first step in the computation was to convert the natural values of peak discharge to logarithms. The mean and the standard deviation for each station array were then computed. It was necessary to adjust the computed mean and the standard deviation for those stations whose peak-discharge records were incomplete for the 28-year base period, 1932–59. The adjustment was accomplished by first correlating concurrent records for each short-term station and a nearby long-term station and then applying equations 3 and 4. At this stage of the computations, the base-period mean (M) and standard deviation (S) were available for all 13 stations.

The values of M and S, still in logarithmic units, were next regionalized by correlation with the hydrologic parameters listed in table 5. In the Coast Ranges region, there was a definite relation between M and drainage area (fig. 15), but S was apparently independent of any measured hydrologic parameters. It was expected that S might vary inversely with mean annual basinwide precipitation, but no correlation was evident. An average value of $S=0.235$ was therefore used for the Coast Ranges. In the Klamath Mountains, M was found to vary with both drainage area and mean annual basinwide precipitation. No graph has been provided to illustrate the relation, which is expressed by the equation:

$$M=-2.837+0.978A+2.421P, \tag{22}$$

where

M is the base-period mean of the logarithms of annual peak discharges,

A is the logarithm of the drainage area, and

P is the logarithm of the mean annual basinwide precipitation.

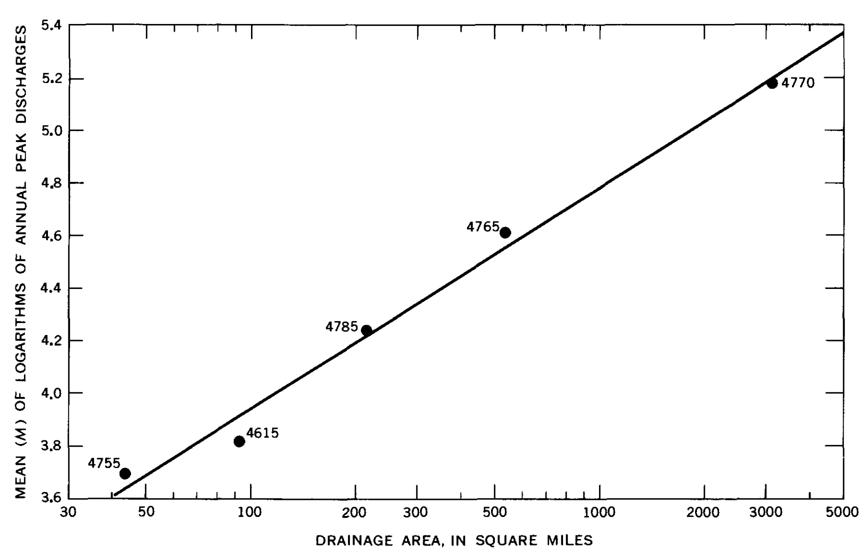


FIGURE 15.—Relation of the mean of the logarithms of annual peak discharges to drainage area in the Coast Ranges.

Values of S in the Klamath Mountains were closely related to mean annual basinwide precipitation as shown graphically in figure 16. By use of these relations, an isohyetal map of mean annual precipitation, and equation 5, a flood-frequency curve can be computed for any site in north coastal California.

By use of the relations just mentioned and equations 12 and 13, values of Q_{50} and Q_{100}, in logarithms, were computed for drainage areas corresponding in size and mean annual basinwide precipitation to those of the 13 long-term gaging stations in the region. The antilogarithms of Q_{50} and Q_{100} have been entered in column 8 of tables 6 and 7, respectively, for later comparison.

EXTREME-VALUE PROBABILITY DISTRIBUTION OR GUMBEL METHOD

A regional flood-frequency study of the north coastal region of California was also made by fitting an extreme-value probability distribution to the base data for the 13 long-term gaging stations in the region. In this analysis natural values of the discharges were used, not their logarithms. The mean and the standard deviation for each station array were computed, and these two statistics were adjusted for those stations whose peak-discharge records were incomplete for the 28-year base period, 1932–59. To make this adjustment, concurrent records for each short-term station and a nearby long-term station were first correlated, and then equations 3 and 4 were applied. (Note that natural values, and not logarithms, are now used in equations 3 and 4.) After making the adjustment, the base-period mean (M) and standard deviation (S) were available for all 13 stations.

The values of M and S were next regionalized by correlation with the hydrologic parameters listed in table 5. In the Coast Ranges region, both M and S were related to drainage area (fig. 17). In the Klamath Mountains both M and S were related to drainage area and mean annual basinwide precipitation. No graphs have been provided to illustrate the two multiple relations, which are expressed by the equations:

$$M=0.0044A^{0.944}P^{2.164} \tag{23}$$

$$S=0.022A^{0.891}P^{1.861}, \tag{24}$$

where

M is the base-period mean of annual peak discharges,

S is the base-period standard deviation of annual peak discharges,

A is the drainage area, and

P is the mean annual basinwide precipitation.

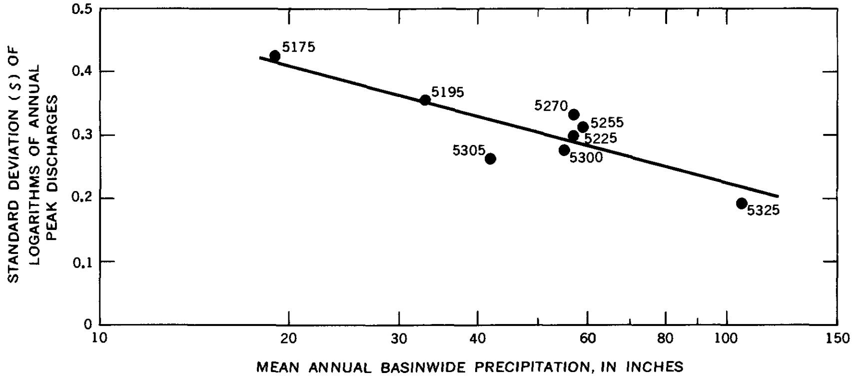


FIGURE 16.—Relation of the standard deviation of the logarithms of annual peak discharges to mean annual basinwide precipitation in the Klamath Mountains.

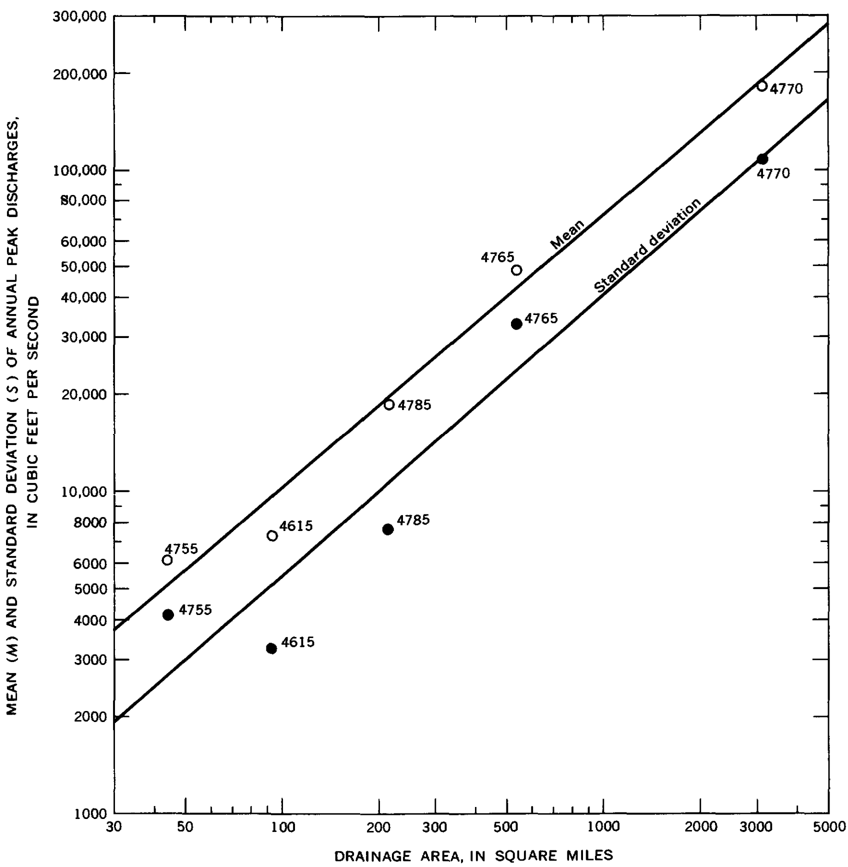


FIGURE 17.—Relation of the mean and the standard deviation of annual peak discharges to drainage area in the Coast Ranges.

By use of the relations expressed in figure 17 and in equations 23, 24, and 6, a flood-frequency curve can be computed for any site in north coastal California.

By use of the foregoing relations and equations 14 and 15, values of Q_{50} and Q_{100} were computed for drainage areas corresponding in size and mean annual basinwide precipitation to those of the 13 long-term gaging stations in the region. These values are listed in column 10 of tables 6 and 7, respectively, for later comparison.

PEARSON TYPE III DISTRIBUTION

A fifth flood-frequency analysis was made for the north coastal region of California by fitting a Pearson type III distribution to the peak-discharge data for the 13 long-term gaging stations in the region. The computation and regionalization of the base period (M) and standard deviation (S) for this study are identical with those for the Gumbel method of analysis. Consequently, the graphs for the Coast Ranges (fig. 17) and equations 23 and 24 for the Klamath Mountains were applicable for this analysis. The next step was to determine the coefficient of skew for each station. This was done by first using equation 7 to compute the annual peak discharges needed to complete the 28-year array at each short-term station, and then applying equation 8 to compute the coefficient of skew (g) for each station. For the streams in the Coast Ranges, the individual values of g ranged from 0.84 to 2.04; for the streams in the Klamath Mountains they ranged from 1.07 to 3.00. In neither region did g correlate with the selected hydrologic parameters. Consequently, the average value of g in each region was assumed to be the regional value of that statistical parameter. Under this assumption the regional value of g for the Coast Ranges is 1.46, and for the Klamath Mountains, 1.83.

By use of the regional relations for M and S as indicated by curves given in figure 17 and equations 23 and 24, the appropriate regional value of g, table 1, and equation 9, a flood-frequency curve can be computed for any site in the region, whether gaged or ungaged. Values of Q_{50} and Q_{100} were computed by use of the relations and table just mentioned for drainage areas corresponding in size and mean annual basinwide precipitation to those of the 13 long-term gaging stations in the region. These values have been entered in column 12 of tables 6 and 7, respectively, for later comparison with the results obtained by the other methods of analysis.

GAMMA DISTRIBUTION

The gamma distribution was also used to analyze the flood magnitude-frequency relation for streams in north coastal California. A prerequisite to the use of this method was a determination of the base-period arithmetic mean (M) and geometric mean (Mg) for each of the 13 long-term stations. Values of M were available from the previous use of the Gumbel-method analysis, and the logarithms of Mg were available from the previous use of the logarithmic normal distribution. The logarithms of Mg were then converted to natural values. The logarithm of the ratio M/Mg was next computed for each station, and the curve given in figure 2 was then used to obtain corresponding values of C. The next step was to regionalize the values of M and C. This had been done for M in the Gumbel method of analysis and the results are found in figure 17 for the Coast Ranges and in equation 23 for the Klamath Mountains. Individual values of C for the five stations in the Coast Ranges ranged from 2.61 to 7.35, but showed no correlation with hydrologic parameters. The median value of C, 3.40, was therefore assumed to be the regional value of C for this area. For streams in the Klamath Mountains, the individual values of C correlated with mean annual basinwide precipitation (fig. 18).

The regional values of M and C can be used with a table of chi-square to compute a flood-frequency curve for any site in the region, whether gaged or ungaged. Values of Q_{50} and Q_{100} were computed by use of the curves given in figures 17 and 18, equation 23, a regional C value of 3.40 for the Coast Ranges, and a chi-square table for drainage areas corresponding in size and mean annual basinwide precipita-

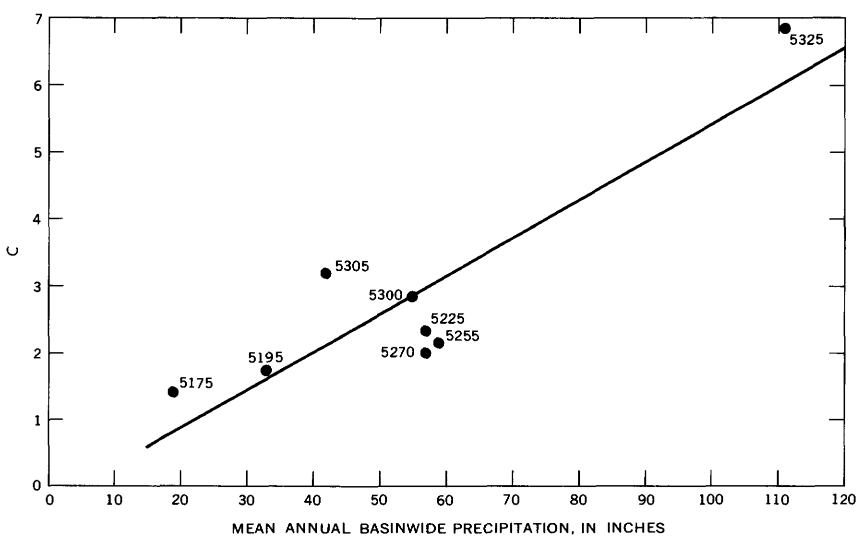


FIGURE 18.—Relation of C to mean annual basinwide precipitation in the Klamath Mountains (gamma distribution).

tion to those of the 13 long-term gaging stations in the region. These values are listed in column 14 of tables 6 and 7, respectively, for later comparison with the results obtained by the other five methods of analysis.

DISCUSSION OF RESULTS OF THE ANALYSES

Because of the short period covered by most records of peak discharge in California, the true values of the 50-year flood (Q_{50}) and the 100-year flood (Q_{100}) are not known. Although the peak discharges of the greatest and second greatest floods in the past 100 years can often be deduced within fairly close limits by the use of qualitative historical information, utilization of this information provides, at best, only a single 100-year sample of flood events. This sample may or may not be representative of long-term flood activity. Furthermore, in each of the two California regions studied, the greatest flood in the past 100 years at each station was caused by a single meteorological event. This was generally true, also, of the second greatest flood in the past 100 years. Thus, all stations experienced the same major flood events and, therefore, regionalization of the data has had little effect in reducing the sampling error that results from having a nonrepresentative sample of floods at a gaging station. We can do nothing about the sampling error, however, and must proceed on the assumption that our sampling of flood events adequately represents the norm of flood activity and that our values of Q_{50} and Q_{100} for each station, as deduced with the help of historic information, have only random scatter from their true values.

Tables 3 and 4 present a summary of values of Q_{50} and Q_{100}, respectively, for streams in the San Diego area, obtained by various methods of flood-frequency analysis. Similar tabulations for north coastal California are given in tables 6 and 7. Because all methods attempt to fit some relation to the single flood array that is available for each station in a region, it is presumed that the method that fits the observed data best is the most reliable one for use. In this study the values of Q_{50} and Q_{100} obtained from the graphically-derived individual station flood-frequency curves (col. 3 in tables 3, 4, 6, and 7) are considered the standard for judging the reliability of the various methods of analysis. Note that these values of Q_{50} and Q_{100} in column 3 are more than mere extrapolations; they have been determined from a careful assessment of qualitative historical information. Tables 3, 4, 6, and 7 also list the differences, in percent, between the values of Q_{50} and Q_{100} as determined by each of the six methods of regional flood-frequency analysis, and the values of Q_{50} and Q_{100} in column 3. These differences are summarized in table 8. Ideally, split-sample testing should have been

TABLE 8.—*Summary of differences between discharges determined from graphically derived flood-frequency curves and discharges computed by various methods of regional flood-frequency analysis*

Range of differences (percent)	Number of sites with differences in indicated range											
	Q_{50}						Q_{100}					
	Index-flood method	Multiple correlation	Logarithmic normal distribution	Extreme-value probability distribution	Pearson type III distribution	Gamma distribution	Index-flood method	Multiple correlation	Logarithmic normal distribution	Extreme-value probability distribution	Pearson type III distribution	Gamma distribution
San Diego area												
+(151–222)									5			
+(126–150)			1						0			
+(101–125)			1						2			
+(81–100)			1						3			
+(61–80)			4				1		1			
+(41–60)	2	1	3				1	1	4			
+(21–40)	2	2	2		1		4	1	1		1	
	—	—	—	—	—	—	—	—	—	—	—	—
+(1–20)	5	4	3	1	1		2	7	2		1	
0	0	3	0	0	0		0	2			0	
−(1–20)	5	5	3	2	6	2	6	4		1	4	1
	—	—	—	—	—	—	—	—	—	—	—	—
−(21–40)	4	3		5	8	6	4	3		6	4	5
−(41–60)				9	2	9				8	8	8
−(61–80)				1		1				3		4
Total	18	18	18	18	18	18	18	18	18	18	18	18

North coastal area

+(61–80)	---	---	---	---	---	---	---	---	1	---	---	---
+(41–60)	---	---	2	---	---	---	---	---	2	---	---	---
+(21–40)	1	4	3	---	2	---	1	4	2	1	2	---
+(1–20)	7	3	3	4	4	2	7	3	3	5	5	2
0	0	1	0	0	1	1	0	1	0	0	0	0
−(1–20)	2	2	3	5	2	1	2	1	2	2	2	2
−(21–40)	2	3	2	3	4	5	2	4	3	4	3	4
−(41–60)	1	---	---	1	---	4	1	---	---	1	1	5
Total	13	13	13	13	13	13	13	13	13	13	13	13

used in comparing computed and recorded values of Q_{50} and Q_{100}. In split-sample testing, half the stations would be used to derive the relations necessary to compute Q_{50} and Q_{100}, and the remaining stations would be used to compare computed and recorded values. Unfortunately, there were too few stations to permit split-sample testing, and consequently all stations were used both in the derivation of the relations and in the comparison of Q_{50} and Q_{100}.

It should be pointed out that the percentage differences listed in table 8 do not, in themselves, give a complete picture of the relative reliability of the various methods; another factor is the number of degrees of freedom lost in each analysis. The details of this loss will not be discussed. Suffice it to say that the index-flood method, as used in the analysis for north coastal California, has more lost degrees of freedom than the other methods because of the numerous areal subdivisions that were established for the region. Consequently, the results obtained in north coastal California by this method have somewhat less statistical significance than those obtained by the other five methods.

Table 8 shows that all methods of analysis gave better results for north coastal California than for the San Diego area. This is to be expected because the San Diego area, being generally subhumid, has highly variable streamflow, and it is almost axiomatic that the more variable the streamflow, the more difficult it is to make a reliable determination of the flood magnitude-frequency relation. The difficulty is compounded when the period of record is short. A comparative measure of the variability of an array of data is the coefficient of variation, which is defined as the ratio of the standard deviation of the array to its mean, or S/M. The greater variability of the San Diego area streams is shown by the fact that the coefficient of variation for these streams ranged from 1.5 to 2.8, whereas in the north coastal area the range was from 0.4 to 1.1.

Table 8 shows that for the San Diego area the two empirical methods—the index-flood method and multiple correlation—gave results that are in much closer agreement with the "standard" values of Q_{50} and Q_{100} than are the values obtained by any of the four methods using statistical distributions. This agreement is not surprising because the two empirical methods are the only ones that made use of the qualitative historical data that were available, and it is on these data that the standard values of Q_{50} and Q_{100} are based. In effect, the period of observed data for the two empirical methods was 100 years or more, as opposed to the much shorter period of observed data used with the other four methods. For the north coastal area, the difference in the quality of results between the two empirical methods and the methods using statistical distributions is less pronounced. The gamma distribution gave the poorest results, but of the other five methods, no single method was outstandingly superior. The implication is that in a humid area where all streams are subject to the same general storms, most methods of analysis will give satisfactory results.

INDEX-FLOOD AND MULTIPLE-CORRELATION METHODS

The multiple-correlation method is superior to the index-flood method, particularly in subhumid areas. Not only did the multiple-correlation method give generally better results in this study, as indicated in table 8, but it has a more rational basis than the index-flood method. The weakness of the index-flood method is that it assumes that the slopes of the flood-frequency curves for the streams in a given region vary randomly from some median value. The Geological Survery test for homogeneity of slope in a region has some serious shortcomings and often indicates homogeneity of slope where none exists (Benson, 1962, p. 21). Actually, there are many physiographic and climatologic factors that affect the slope of the flood-frequency curve, and it is often a matter of fortuity that the curves for all streams in a region have similar slopes. In several previous studies the delineation of the boundaries of regions of homogeneous slope, based on the homogeneity test, has resulted in maps that resemble political gerrymanders. The index-flood method can be used to advantage, however, for a region where there are insufficient stations with 10 or more years of record to permit a significant multiple correlation to be made. In addition, those records that are shorter than 10 years are helpful in defining the regional relation for the mean annual flood ($Q_{2.33}$).

STATISTICAL DISTRIBUTIONS

We turn our attention now to the four methods of analysis that are based on statistical distributions—logarithmic normal, extreme-value probability, Pearson type III, and gamma. As already mentioned, these methods are empirical in the sense that one of the four distributions must be arbitrarily selected for use. However, once a distribution has been selected, the analysis becomes strictly objective. Because these methods did not make use of the available historical information, the results obtained from their application were not as satisfactory as those obtained by use of either the index-flood or multiple-correlation methods. This was particularly true in the subhumid San Diego area. If the high-water ends of these distributions had been later modified on the basis of the historical information, as is often done, the results would have been more satisfactory, but there would then have been little point in using one of the theoretical distributions. The all-important high-water end of the flood magnitude-frequency relation would no longer be controlled by the statistics of the selected distribution, but would depend on the judgment of the analyst in interpreting the historical information and assigning probabilities to the peak discharges. Table 8 shows that the logarithmic normal distribution gave results that were generally high; the other three distributions gave results that tended to be low. The logarithmic normal distribution also had a greater spread of differences from the standard values of Q_{50} and Q_{100} than did the other three distributions. It is not to be inferred, however, that this is a general rule. It is of interest that the flood-frequency handbook used by the U.S. Army, Corps of Engineers, which recommends the logarithmic normal distribution method, cautions against its use in subhumid areas (Beard, 1962, p. 22).

Tables 3, 4, 6, and 7 show that the four statistical distributions, when applied to the same peak-discharge data, gave widely differing results for individual stations. These distributions have all been extensively used elsewhere and each has the support of reputable statistician-hydrologists as being the distribution that best describes the occurrence of flood events. Obviously all methods cannot be best, and the inescapable conclusion is either that the true distribution that describes the occurrence of flood events is not known or that no single distribution is best for all of the many widely varying hydrologic conditions found in a country as large as the United States. In the United States most of the long-term discharge records are for large streams draining humid areas, and it is these records that have been used in the past to test the adequacy of the various distributions. In this study we have seen that in a humid area all methods give results that are generally satisfactory. Therefore, it is not surprising that each distribution has strong proponents among hydrologists. Examination of the various distributions shows that within the range of probabilities usually used in flood-frequency studies, the extreme-value probability distribution may be considered a special case of the Pearson type III distribution, one in which the coefficient of skew is 1.14. The gamma distribution is likewise a special case of the Pearson type III distribution, one in which the locus parameter is zero.

From the preceding paragraph, it would appear that the most desirable distribution for use would be the one that is most flexible and can therefore fit the peak-flow data closest. Of the four distributions studied, the Pearson type III distribution is the least rigid because the Pearson type III distribution uses three statistical parameters, whereas the other three distributions use two parameters, with the third parameter either constant or determined by one or both of the others. We know too little of the laws governing the occurrence of flood events to hazard a guess concerning the value of the third parameter, and it therefore seems reasonable to use a general or flexible distribution, such as the Pearson type III, rather than some rigid distribution, such as one of the other three, which has a built-in restriction on the value of the third parameter.

If we exclude the two empirical methods and confine our comparison of results to the four statistical distributions, we should concern ourselves only with the values of Q_{50} and not Q_{100}. In this study, the values obtained for Q_{50} result primarily from the distribution of peak discharges during the base periods; they are not seriously affected by the consideration of qualitative historical data, as are the values of Q_{100}. Examination of the percent differences for Q_{50} in table 8 shows that the Pearson type III distribution fitted the array of observed data better than any of the other three statistical distributions.

The principal objection to the use of Pearson type III distribution has been the fact that the skew statistic, which is used in the distribution, has a large standard error when there are relatively few items in the array of peak discharges. It has been claimed that under these circumstances the coefficient of skew has little or no significance. Whether or not the coefficient of skew computed for an isolated station is statistically significant may be debatable, but it seems reasonable to accept as significant a regionalized value of the coefficient based on numerous station records, even though the records may not be entirely independent.

Because regionalization of the statistics for the individual station arrays might have obscured the relative abilities of the four distributions to fit the station data, the individual statistics for each station array were used to compute individual station flood-frequency curves. The computed value of Q_{50} from each of these curves was compared with the "standard" values of Q_{50} listed in column 3 of tables 3, 4, 6, and 7. The comparison is summarized in table 9. The Pearson type III distribution gave results that were superior to those obtained by use of the other three distributions.

Although this comparison study of flood-frequency methods was based on small samples—18 stations in southwest California and 13 stations in northwest California—the results and conclusions appear to be meaningful because they can be explained rationally.

TABLE 9.—*Summary of differences between* Q_{50} *determined from graphically derived flood-frequency curves and* Q_{50} *computed from various distributions using statistical parameters from individual station arrays*

Range of differences (percent)	Number of sites with differences in indicated range			
	Logarithmic normal distribution	Extreme-value probability distribution	Pearson type III distribution	Gamma distribution
San Diego area				
+(121–373)	3			
+(101–120)	2		1	
+(81–100)	0		0	
+(61–80)	2	1	0	
+(41–60)	4	0	0	1
+(21–40)	1	0	0	0
+(1–20)	0	0	1	0
0	0	0	0	0
−(1–20)	4	1	6	2
−(21–40)	0	8	7	6
−(41–60)	1	6	1	7
−(61–84)	1	2	2	2
Total	18	18	18	18
North coastal area				
+(31–34)	1			
+(21–30)	0			
+(11–20)	3			
+(1–10)	5	1	8	
0	0	0	0	0
−(1–10)	3	11	4	5
−(11–20)	1	1	0	7
−(21–26)			1	1
Total	13	13	13	13

SUMMARY AND CONCLUSIONS

Regional flood-frequency studies were made by six different methods for the subhumid San Diego area in southwestern California and for the humid coastal area in northwestern California. A 55-year base period was used for 18 peak-discharge records in the San Diego area; a 28-year base period was used for 13 peak discharge records in north coastal California. All flood-frequency curves were extrapolated to the discharges corresponding to the 50-year recurrence interval (Q_{50}) and to the 100-year recurrence interval (Q_{100}). Two of the six methods of analysis—the index-flood and multiple-correlation methods—are to some extent empirical and permit the analyst to use whatever qualitative or historical information is available for extrapolating the flood-frequency relation. The other four methods are based on statistical distributions—the logarithmic normal, extreme-value probability, Pearson type III, and gamma distributions. These four methods are empirical in the sense that one of the distributions is arbitrarily selected for use; but once selected, the analysis becomes strictly objective and the extrapolation is automatically made from a determination of the statistics of the array of peak discharges.

There is a lack of agreement among hydrologists as to which method or which distribution is the most reliable for defining the flood magnitude-frequency relation at a site. The purpose of this study was to compare the results obtained by applying each of the six methods of analysis to the basic data and to appraise the relative reliability of the methods. The magnitude of the greatest and second greatest flood peaks in the past 100 years in the two study areas were generally known within reasonable limits of accuracy, and with this knowledge it was possible to derive values of Q_{50} and Q_{100} for each gaging station that are representative of flood activity during the past century. These values of Q_{50} and Q_{100} were used as standards for the comparison and appraisal of the methods of analysis.

It was concluded that all methods of analysis give better results in a humid region, such as north coastal California, than in a subhumid region, such as the San Diego area, because streamflow is usually less variable in a humid region.

A decision as to the preferred method of flood-frequency analysis depends on whether or not historical data, either quantitative or qualitative, are available concerning the magnitude of floods that occurred in the years prior to the collection of streamflow records. If such information is available, the empirical methods—the index-flood and multiple-correlation methods—are superior to any of the four methods using statistical distributions owing to the fact that only the empirical methods use historical data, and use of these data gives a longer time base for the analysis. In this study, our knowledge of the magnitude of the greatest and second greatest flood peaks in the past 100 years extended the period of observed data, in effect, to 100 years or more for the empirical methods. This is a much longer period than the base periods that were used with the other four methods of analysis. Of the two empirical methods, the multiple-regression method is superior because it has a much more rational basis than the index-flood method and in addition gives better results.

Where the peak-discharge data are limited entirely to the period during which streamflow records were collected (no historical data available), a method based on the distribution of the array of peak flows is preferred because of its greater objectivity. Of the four distributions tested, the Pearson type III is the most desirable. It is more flexible than the other three and will generally fit the peak-discharge data best because the Pearson type III distribution is a three-parameter distribution with no built-in restriction on the value of the third, or skew, parameter. The other distributions are two-parameter distributions with an implied constant value for the third parameter. Objection to the Pearson type III distribution has been based on the large standard error of the coefficient of skew, but this objection can be overcome by using a regional value of the coefficient based on numerous station records.

Although this comparison study of flood-frequency methods was based on small samples from only one part of the United States, the results and conclusions appear to be meaningful because they can be explained rationally.

REFERENCES CITED

Alexander, G. N., 1962, The use of the gamma distribution in estimating regulated output from storages: Australian Inst. of Engineers, Civil Engineers Trans., v. CE 4, no. 1, p. 29–34.

Arkin, Herbert and Colton, R. R., 1950, Tables for statisticians: New York, Barnes & Noble, 152 p.

Beard, L. R., 1962, Statistical methods in hydrology: U.S. Army Corps of Engineers, Civil Works Investigations Project CW–151, 62 p.

Benson, M. A., 1962, Evolution of methods for evaluating the occurrence of floods: U.S. Geol. Survey Water-Supply Paper 1580–A, 30 p.

California Department of Water Resources, 1963, Flood hazard investigation, San Diego County: Bull. 112, app. A, 78 p.

Dalrymple, Tate, 1960, Flood-frequency analyses: U.S. Geol. Survey Water-Supply Paper 1543–A, 80 p.

Elderton, W. P., 1953, Frequency curves and correlation; Washington, Harren Press, 272 p.

Ezekiel, Mordecai and Fox, K. A., 1959, Methods of correlation and regression analysis: New York, John Wiley & Sons, 548 p.

Fenneman, N. M.. 1931, Physiography of western United States: New York, McGraw-Hill Book Co., 534 p.

——— 1938, Physiography of eastern United States: New York, McGraw-Hill Book Co., 714 p.

Gumbel, E. J., 1958, Statistics of extremes: New York, Columbia Univ. Press, 375 p.

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

Potter, W. D., 1961, Peak rates of runoff from small watersheds: U.S. Bur. Public Roads, Hydraulic Design Series No. 2, 35 p.

Rantz, S. E., 1964, Surface-water hydrology of coastal basins of northern California: U.S. Geol. Survey Water-Supply Paper 1758, 77 p.

Thom, H. C. S., 1958, A note on the gamma distribution: Monthly Weather Rev., v. 86, no. 4, p. 117–122.

O

Flood Hydrology

GEOLOGICAL SURVEY WATER-SUPPLY PAPER 1580

This volume was published as separate chapters A–E

UNITED STATES DEPARTMENT OF THE INTERIOR

STEWART L. UDALL, *Secretary*

GEOLOGICAL SURVEY

Thomas B. Nolan, *Director*

CONTENTS

[Letters designate the separately published chapters]