Techniques of Water-Resources Investigations of the United States Geological Survey

Chapter A2

FREQUENCY CURVES

By H. C. Riggs

Book 4

HYDROLOGIC ANALYSIS AND INTERPRETATION

DEPARTMENT OF THE INTERIOR
DONALD PAUL HODEL, Secretary

U.S. GEOLOGICAL SURVEY
Dallas L. Peck, Director

First printing 1968
Second printing 1969
Third printing 1978
Fourth printing 1989

UNITED STATES GOVERNMENT PRINTING OFFICE : 1968

For sale by the
Books and Open-File Reports Section
U.S. Geological Survey
Federal Center, Box 25425
Denver, CO 80225

PREFACE

The series of manuals on techniques describes procedures for planning and executing specialized work in water-resources investigations. The material is grouped under major headings called books and further subdivided into sections and chapters; section A of Book 4 is on statistical analysis.

The unit of publication, the chapter, is limited to a narrow field of subject matter. This format permits flexibility in revision and publication as the need arises.

Provisional drafts of chapters are distributed to field offices of the U.S. Geological Survey for their use. These drafts are subject to revision because of experience in use or because of advancement in knowledge, techniques, or equipment. After the technique described in a chapter is sufficiently developed, the chapter is published and is sold by the U.S. Geological Survey Books and Open-File Reports Section, Federal Center, Box 25425, Denver, CO 80225 (authorized agent of Superintendent of Documents, Government Printing Office).

TECHNIQUES OF WATER-RESOURCES INVESTIGATIONS OF THE UNITED STATES GEOLOGICAL SURVEY

The U.S. Geological Survey publishes a series of manuals describing procedures for planning and conducting specialized work in water-resources investigations. The manuals published to date are listed below and may be ordered by mail from the **U.S. Geological Survey, Books and Open-File Reports, Federal Center, Box 25425, Denver, Colorado 80225** (an authorized agent of the Superintendent of Documents, Government Printing Office).

Prepayment is required. Remittance should be sent by check or money order payable to U.S. Geological Survey. Prices are not included in the listing below as they are subject to change. **Current prices can be obtained** by writing to the USGS, Books and Open File Reports. Prices include cost of domestic surface transportation. For transmittal outside the U.S.A. (except to Canada and Mexico) a surcharge of 25 percent of the net bill should be included to cover surface transportation. When ordering any of these publications, please give the title, book number, chapter number, and "U.S. Geological Survey Techniques of Water-Resources Investigations."

TWI 1-D1. Water temperature—influential factors, field measurement, and data presentation, by H.H. Stevens, Jr., J.F. Ficke, and G.F. Smoot. 1975. 65 pages.
TWI 1-D2. Guidelines for collection and field analysis of ground-water samples for selected unstable constituents, by W.W. Wood. 1976. 24 pages.
TWI 2-D1. Application of surface geophysics to ground water investigations, by A.A.R. Zohdy, G.P. Eaton, and D.R. Mabey. 1974. 116 pages.
TWI 2-D2. Application of seismic-refraction techniques to hydrologic studies, by F.P. Haeni. 1988. 86 p.
TWI 2-E1. Application of borehole geophysics to water- resources investigations, by W.S. Keys and L.M. MacCary. 1971. 126 pages.
TWI 3-A1. General field and office procedures for indirect discharge measurement, by M.A. Benson and Tate Dalrymple. 1967. 30 pages.
TWI 3-A2. Measurement of peak discharge by the slope-area method, by Tate Dalrymple and M.A. Benson. 1967. 12 pages.
TWI 3-A3. Measurement of peak discharge at culverts by indirect methods, by G.L. Bodhaine. 1968. 60 pages.
TWI 3-A4. Measurement of peak discharge at width contractions by indirect methods, by H.F. Matthai. 1967. 44 pages.
TWI 3-A5. Measurement of peak discharge at dams by indirect methods, by Harry Hulsing. 1967. 29 pages.
TWI 3-A6. General procedure for gaging streams, by R.W. Carter and Jacob Davidian. 1968. 13 pages.
TWI 3-A7. Stage measurements at gaging stations, by T.J. Buchanan and W.P. Somers. 1968. 28 pages.
TWI 3-A8. Discharge measurements at gaging stations, by T.J. Buchanan and W.P. Somers. 1969. 65 pages.
TWI 3-A9. Measurement of time of travel and dispersion in streams by dye tracing, by E.F. Hubbard, F.A. Kilpatrick, L.A. Martens, and J.F. Wilson, Jr. 1982. 44 pages.
TWI 3-A10. Discharge ratings at gaging stations, by E.J. Kennedy. 1984. 59 pages.
TWI 3-A11. Measurement of discharge by moving-boat method, by G.F. Smoot and C.C. Novak. 1969. 22 pages.
TWI 3-A12. Fluorometric procedures for dye tracing, Revised, by James F. Wilson, Jr., Ernest D. Cobb, and Frederick A. Kilpatrick. 1986. 41 pages.
TWI 3-A13. Computation of continuous records of streamflow, by Edward J. Kennedy. 1983. 53 pages.
TWI 3-A14. Use of flumes in measuring discharge, by F.A. Kilpatrick, and V.R. Schneider. 1983. 46 pages.
TWI 3-A15. Computation of water-surface profiles in open channels, by Jacob Davidian. 1984. 48 pages.
TWI 3-A16. Measurement of discharge using tracers, by F.A. Kilpatrick and E.D. Cobb. 1985. 52 pages.
TWI 3-A17. Acoustic velocity meter systems, by Antonius Laenen. 1985. 38 pages.
TWI 3-B1. Aquifer-test design, observation, and data analysis, by R.W. Stallman. 1971. 26 pages.
TWI 3-B2.[1] Introduction to ground-water hydraulics, a programmed text for self-instruction, by G.D. Bennett. 1976. 172 pages.
TWI 3-B3. Type curves for selected problems of flow to wells in confined aquifers, by J.E. Reed. 1980. 106 pages.
TWI 3-B5. Definition of boundary and initial conditions in the analysis of saturated ground-water flow systems—an introduction, by O. Lehn Franke, Thomas E. Reilly, and Gordon D. Bennett. 1987. 15 pages.
TWI 3-B6. The principle of superposition and its application in ground-water hydraulics, by Thomas E. Reilly, O. Lehn Franke, and Gordon D. Bennett. 1987. 28 pages.
TWI 3-C1. Fluvial sediment concepts, by H.P. Guy. 1970. 55 pages.
TWI 3-C2. Field methods of measurement of fluvial sediment, by H.P. Guy and V.W. Norman. 1970. 59 pages.
TWI 3-C3. Computation of fluvial-sediment discharge, by George Porterfield. 1972. 66 pages.
TWI 4-A1. Some statistical tools in hydrology, by H.C. Riggs. 1968. 39 pages.
TWI 4-A2. Frequency curves, by H.C. Riggs, 1968. 15 pages.
TWI 4-B1. Low-flow investigations, by H.C. Riggs. 1972. 18 pages.
TWI 4-B2. Storage analyses for water supply, by H.C. Riggs and C.H. Hardison. 1973. 20 pages.
TWI 4-B3. Regional analyses of streamflow characteristics, by H.C. Riggs. 1973. 15 pages.
TWI 4-D1. Computation of rate and volume of stream depletion by wells, by C.T. Jenkins. 1970. 17 pages.
TWI 5-A1. Methods for determination of inorganic substances in water and fluvial sediments, by M.W. Skougstad and others, editors. 1979. 626 pages.

[1]Spanish translation also available.

TWI 5-A2. Determination of minor elements in water by emission spectroscopy, by P.R. Barnett and E.C. Mallory, Jr. 1971. 31 pages.

TWI 5-A3. Methods for the determination of organic substances in water and fluvial sediments, edited by R.L. Wershaw, M.J. Fishman, R.R. Grabbe, and L.E. Lowe. 1987. 80 pages. This manual is a revision of "Methods for Analysis of Organic Substances in Water" by Donald F. Goerlitz and Eugene Brown, Book 5, Chapter A3, published in 1972.

TWI 5-A4. Methods for collection and analysis of aquatic biological and microbiological samples, edited by P.E. Greeson, T.A. Ehlke, G.A. Irwin, B.W. Lium, and K.V. Slack. 1977. 332 pages.

TWI 5-A5. Methods for determination of radioactive substances in water and fluvial sediments, by L.L. Thatcher, V.J. Janzer, and K.W. Edwards. 1977. 95 pages.

TWI 5-A6. Quality assurance practices for the chemical and biological analyses of water and fluvial sediments, by L.C. Friedman and D.E. Erdmann. 1982. 181 pages.

TWI 5-C1. Laboratory theory and methods for sediment analysis, by H.P. Guy. 1969. 58 pages.

TWI 6-A1. A modular three-dimensional finite-difference ground-water flow model, by Michael G. McDonald and Arlen W. Harbaugh. 1988. 586 pages.

TWI 7-C1. Finite difference model for aquifer simulation in two dimensions with results of numerical experiments, by P.C. Trescott, G.F. Pinder, and S.P. Larson. 1976. 116 pages.

TWI 7-C2. Computer model of two-dimensional solute transport and dispersion in ground water, by L.F. Konikow and J.D. Bredehoeft. 1978. 90 pages.

TWI 7-C3. A model for simulation of flow in singular and interconnected channels, by R.W. Schaffranek, R.A. Baltzer, and D.E. Goldberg. 1981. 110 pages.

TWI 8-A1. Methods of measuring water levels in deep wells, by M.S. Garber and F.C. Koopman. 1968. 23 pages.

TWI 8-A2. Installation and service manual for U.S. Geological Survey monometers, by J.D. Craig. 1983. 57 pages.

TWI 8-B2. Calibration and maintenance of vertical-axis type current meters, by G.F. Smoot and C.E. Novak. 1968. 15 pages.

CONTENTS

FIGURES

TABLES

FREQUENCY CURVES

By H. C. Riggs

Abstract

This manual describes graphical and mathematical procedures for preparing frequency curves from samples of hydrologic data. It also discusses the theory of frequency curves, compares advantages of graphical and mathematical fitting, suggests methods of describing graphically defined frequency curves analytically, and emphasizes the correct interpretations of a frequency curve.

Introduction

A frequency curve relates magnitude of a variable to frequency of occurrence. The curve is an estimate of the cumulative distribution of the population of that variable and is prepared from a sample of data.

Frequency curves have many uses in hydrology. Flood-frequency curves are widely used in the design of bridge openings, channel capacities, and roadbed elevations; for flood-plain zoning; and in studies of economics of flood-protection works. Frequency curves of annual low flows are used in design of industrial and domestic water-supply systems, classification of streams as to their potential for waste dilution, definition of the probable amount of water available for supplemental irrigation, and maintenance of certain channel discharges as required by agreement or by law. Frequency curves of annual mean flows are sometimes used in studies of the carryover of annual storage (Beard, 1964).

Frequency curves also provide a means of classifying data for use in subsequent analyses. For example, Benson (1962a) used intensity of rainfall for a given frequency in his regional flood-frequency analysis for New England, and Riggs (1953) used a frequency curve of runoff in excess of assured flow in a forecasting problem. Many other applications have been and can be made.

Cumulative Distributions

Book 4, chapter A1 of the series of Techniques of Water-Resources Investigations (Riggs, 1967) describes the relation of a frequency distribution or probability density curve to its cumulative form. A more detailed examination of this relation helps in understanding the cumulative distribution, or frequency curve. We begin with the two normal distributions shown in figure 1. Their cumulative forms can be expressed as straight lines by use of the special abscissa scale which is derived from the characteristics of the normal distribution. Both distributions have the same median value, 20, and these medians plot at 0.5 probability on the cumulative graph. The variability of a distribution is indicated by the slope of the cumulative distribution; that is, the greater the variability, the greater the slope. The standard deviation is half the difference between magnitudes at probabilities of 0.16 and 0.84 (Dixon and Massey, 1957, table A-4).

Many frequency distributions are nonsymmetrical. For such distributions, the mean, median, and mode have different values which consequently correspond to different probabilities on the cumulative graph. A nonsymmetrical distribution is classified as skewed. Skewness may be shown graphically as right or left; it may be described mathematically by a number, either positive or negative. Two skewed distributions and a symmetrical distribution are shown in figure 2, which also shows the corresponding cumulative distributions (frequency curves).

For a normal, or any symmetrical, distribution the mean and median are the same value. Thus, the value corresponding to the probability of 0.5 on the cumulative frequency curve is the mean as well as the median for such

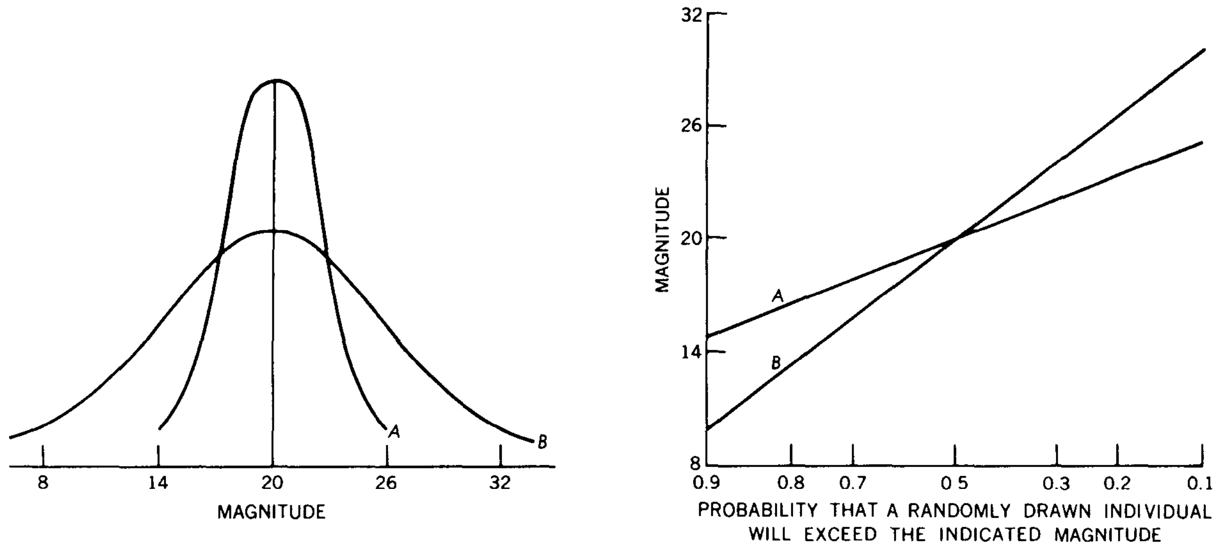

Figure 1.—Two normal distributions and their cumulative forms.

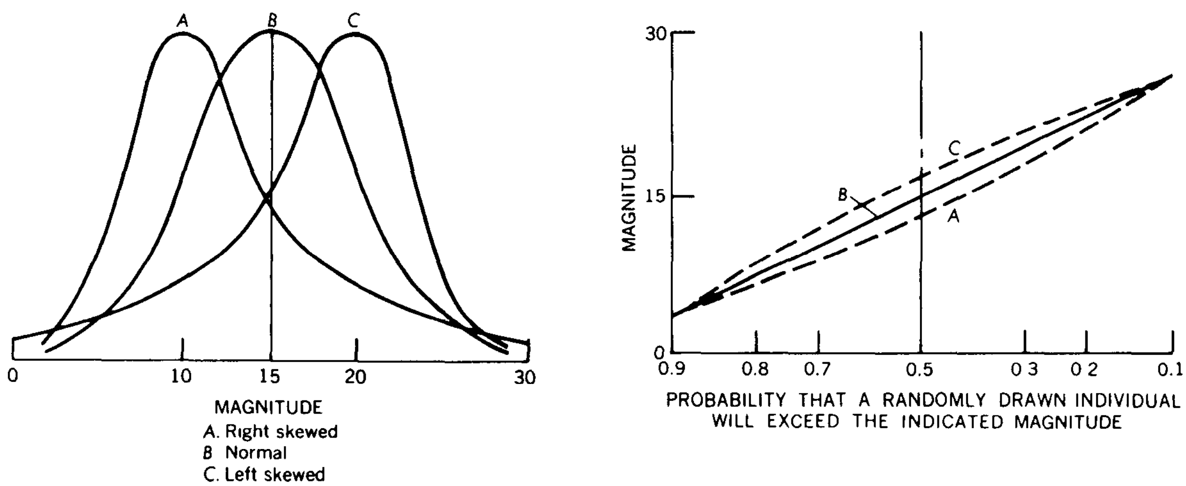

Figure 2.—Normal and skewed distributions and their cumulative forms on a normal-probability plot.

distributions. The relative positions of the mean, median, and mode for skewed distributions are shown in figure 3. Only the median value can be determined from the cumulative plot. The position of the mean with respect to the median on the cumulative plot depends on the degree of skewness, the direction of skewness, and the direction in which the frequency distribution is cumulated. For example, the mean of a particular right (positive)-skewed distribution will be exceeded 43 percent of the time; but 57 percent of the time it will not be exceeded. Thus, if the distribution is cumulated from the high end, the mean is to the right of the median; if cumulated from the low end, the mean is to the left of the median. These relations are reversed for a left-skewed distribution. Figure 4 illustrates the relations. The probability scales of the two plots of figure 4 are different. Each is designed for the particular distribution plotted.

Frequency curves of a time series commonly relate magnitude to recurrence interval or return period instead of to probability of exceedence or nonexceedence. Recurrence interval is the average length of time between exceeden-

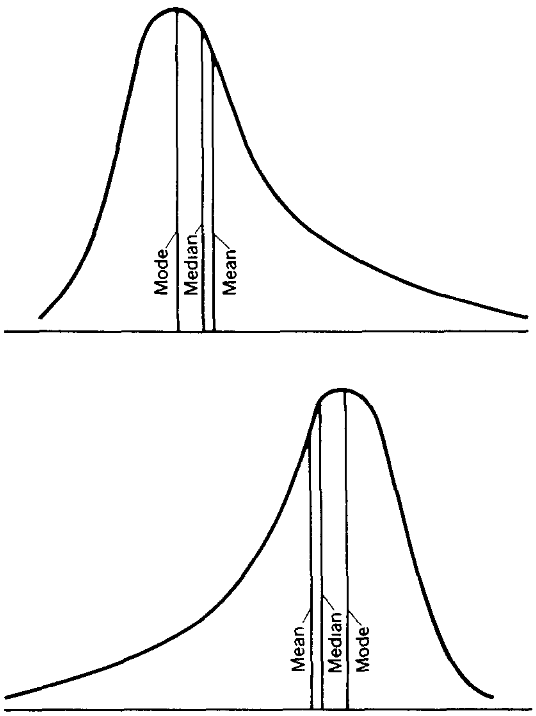

Figure 3.—Relative positions of the mean, median, and mode for right-skewed (upper) and left-skewed (lower) distributions.

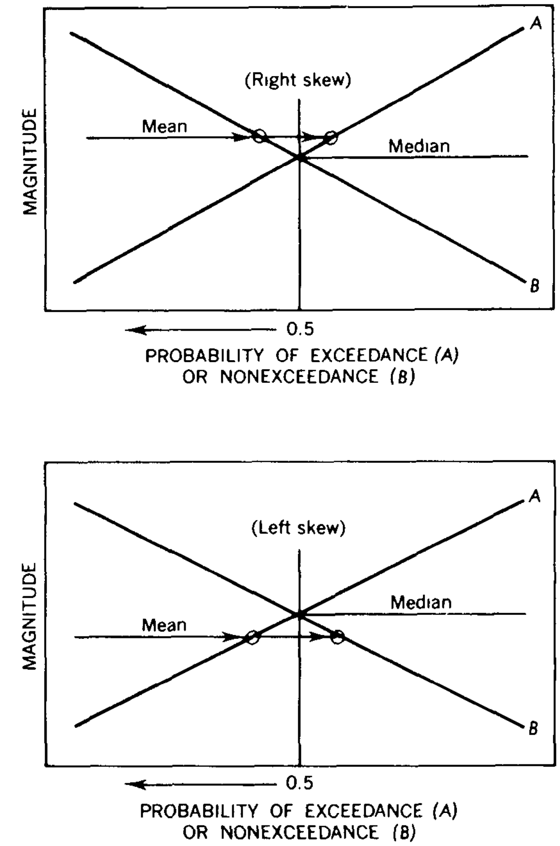

Figure 4.—Frequency curves showing effect of direction of skew and direction of cumulation on position of the mean with respect to the median.

ces, or nonexceedences, of a particular magnitude. It is also defined as the reciprocal of the probability of exceedence (Gumbel, 1954a; Langbein, 1960, p. 48). Recurrence intervals of hydrologic events are usually stated in years and thus are reciprocals of probabilities of exceedence in one year. Further discussion of the meaning of recurrence interval is given in the section on "Interpretation of frequency curves."

Distributions Used in Hydrology

Hydrologists have long sought one theoretical distribution that would describe flood events. If there were such a universal distribution, the observed distributions of flood events at various sites would differ only in the parameters in that universal distribution, and in sampling error. Basin characteristics influence the distribution of floods, so that it seems unlikely that any one theoretical distribution would be generally applicable. It is well established that the distributions of annual minimum flows are highly dependent on basin characteristics (Riggs 1965).

A sample of only 20 or 30 items may define a frequency curve which differs greatly from the population frequency curve. Thus, frequency curves based on small samples may appear very dissimilar, and yet the corresponding population frequency curves may be similar.

As a consequence of the variability of characteristics from basin to basin and of the sampling variability in time, several theoretical distributions are used in hydrology. Characteristics of the more common ones are described below. In addition, graphically defined distributions (those having no known underlying formula) are widely used.

Normal Distribution

The equation of the normal probability density curve is

$$f(X)=(1/\sigma\sqrt{2\pi})\,e^{-(X-\mu)^2/2\sigma^2},$$

where $f(X)$ is the probability-density function, X is the event described, and the parameters of the distribution are the mean, μ, and the variance, σ^2 (commonly reported as σ, the standard deviation). The density curve is bell shaped and symmetrical; therefore the mean and median are the same. Values of X corresponding to various cumulative probabilities are tabulated in most statistics texts. Normal probability plotting paper, available commercially, is designed so that any cumulative normal distribution plots as a straight line on it. The range of the normal distribution is from minus to plus infinity.

Lognormal distribution

Lognormal distribution is a normal distribution of $\ln X$ (Naperian logarithm of X). In terms of X, it is a three-parameter distribution having a range from zero to plus infinity. The statistical parameters of X are given by Chow (1964). In practice, X is plotted on common-logarithmic probability paper, and the parameters given by Chow using Napierian logarithms do not apply. The lognormal distribution can be treated simply as a normal distribution of logarithms, or in a complex manner as a skewed distribution of the untransformed data.

Type I extreme-value distribution (Gumbel)

Type I extreme-value distribution, the first asymptotic distribution (Gumbel, 1958), has two parameters, but it has a fixed skew of 1.139 and therefore is not symmetrical about the mean. Use of this distribution for annual floods was proposed by Gumbel (1941). Powell (1943) prepared the plotting paper based on this distribution; the Geological Survey Form 9–179a is a slight modification of Powell's plot. The mean of the distribution occurs at the 2.33-year recurrence interval when the distribution is cumulated from the upper end. Chow (1954) shows by his table 3 that for practical purposes the extreme-value law is but a special case of the log-probability law.

Type III extreme-value distribution

Type III extreme-value distribution is also called the Weibull distribution after the man who first used it in analysis of strength of materials. Gumbel (1954b) has applied it to drought-frequency analysis. The distribution has three parameters, a lower limit (which may be zero or a finite value greater than zero), a characteristic value which has a recurrence interval of 1.58 years, and a parameter which defines skewness.

Pearson Type III distribution

The Pearson Type III distribution is a flexible distribution in three parameters with a limited range in the left direction and unlimited range to the right. Plotting paper is not available for this distribution because skewness varies. This distribution is commonly fitted to the logarithms of flood magnitudes rather than to the magnitudes themselves because this results in a smaller skew. The Pearson Type III distribution having zero skew is identical to the normal distribution.

The gamma distributions, sometimes used in hydrology, are in effect the Type III curves of Pearson (Mood, 1950, p. 118).

Graphically defined distributions

Distributions defined by graphical means may conform to some theoretical distribution but ordinarily do not.

Mathematical Curve Fitting

Normal distribution

Compute mean and standard deviation of sample. Using these, the detailed characteristics of the distribution can be extracted from a table of the cumulative normal distribution. For example, suppose the mean and standard deviation are computed as 100 and 20, respectively. We assume that these sample parameters, $\overline{X}$ and S, are equal to their respective population parameters, μ and σ. Then the magnitude of the variable at selected levels, call it X_a, can be computed from μ and σ; and the probabilities that a random value of X will be less than these values of X_a are taken from table A–4, page 382, Dixon and Massey (1957). In this table, "area" is equivalent to probability. The last two columns of the following table are computed.

X_*	$P(X<X_*)$	$P(X>X_*)$	Recurrence interval of exceedence
$\mu-2.0\sigma=60$	0.02	0.98	1.02
$\mu-1.5\sigma=70$	.07	.93	1.08
$\mu-1.0\sigma=80$	.16	.84	1.19
$\mu-.8\sigma=84$	.21	.79	1.27
$\mu-.5\sigma=90$	.31	.69	1.45
$\mu-.2\sigma=96$	.42	.58	1.72
$\mu=100$	.50	.50	2.00
$\mu+.2\sigma=104$	.58	.42	2.38
$\mu+.5\sigma=110$	.69	.31	3.22
$\mu+.8\sigma=116$	.79	.21	4.76
$\mu+1.0\sigma=120$	.84	.16	6.25
$\mu+1.5\sigma=130$	.93	.07	14.4
$\mu+2.0\sigma=140$	98	.02	50.0

If the sample is drawn from a time series of annual values, the computed recurrence interval is in years. The same results could have been obtained from a plot on normal probability paper. The mean is plotted at 0.5 probability, the standard deviation is plotted plus and minus from the mean at probabilities of 0.16 and 0.84, respectively, a straight line is drawn through the plotted points, and probabilities at selected levels are read from the line.

Three-parameter distributions

Compute mean, $\overline{X}$, standard deviation, S, and skew coefficient, C_s, by the following equations:

$$\overline{X}=\sum X/N$$

$$S^2=\frac{\sum X^2-(\sum X)^2/N}{N-1}$$

$$C_s=\frac{N^2\sum X^3-3N\sum X\sum X^2+2(\sum X)^3}{N(N-1)(N-2)S^3},$$

where X is the magnitude of an event, and N is the number of events in the sample. These sample parameters are treated as though they were the population parameters in fitting a distribution. These parameters could be substituted in the formula for the distribution to be used, but the distribution cannot be integrated directly. Therefore, the relation between magnitude and probability of exceedence (or nonexceedence) is commonly determined from a table of frequency factors for the chosen distribution and from the general formula

$$X=\overline{X}+KS,$$

where X is the variable, $\overline{X}$ is the mean of the sample, S is the standard deviation of the sample, and K is the frequency factor. For example, in the table under the section on "Normal distribution," the coefficients of σ in the first column are frequency factors, K, for the normal distribution.

Frequency factors for the lognormal distribution are given by Chow (1964, p. 8–26). Recurrence interval is the reciprocal of the probability given in the table. A table by Hazen (1930) has been widely used, but it was developed empirically and contains some values which differ from the theoretical ones. Chow's table shows a definite theoretical relation between the coefficient of variation, C_v, defined as

$$C_v=S/\overline{X}$$

and the coefficient of skew, C_s. Values of C_v and C_s computed from a sample will rarely be related according to theory because the coefficient of skew computed from a few items is notably unreliable. If C_v and C_s define a much different relation than prescribed by theory, the lognormal distribution may provide poor fit to the data. Matalas and Benson (personal commun., 1964) show the standard error of the skew coefficient for N from 4 to 100.

Frequency factors for the Pearson Type III distribution, adapted from a table by Beard (1962), are given in table 1. Chow (1964) shows

Table 1.—Frequency factors for Pearson Type III distributions

[From Beard, 1962]

C_s	Recurrence interval (years)							
	100	20	10	3.33	2	1.43	1.11	1.05
1.0	3.03	1.87	1.34	0.38	−0.16	−0.61	−1.12	−1.31
.8	2.90	1.83	1.34	.42	−.13	−.60	−1.16	−1.38
.6	2.77	1.79	1.33	.45	−.09	−.58	−1.19	−1.45
.4	2.62	1.74	1.32	.48	−.06	−.57	−1.22	−1.51
.2	2.48	1.69	1.30	.51	−.03	−.55	−1.25	−1.58
0.0	2.33	1.64	1.28	.52	0.00	−.52	−1.28	−1.64
−.2	2.18	1.58	1.25	.55	.03	−.51	−1.30	−1.69
−.4	2.03	1.51	1.22	.57	.06	−.48	−1.32	−1.74
−.6	1.88	1.45	1.19	.58	.09	−.45	−1.33	−1.79
−.8	1.74	1.38	1.16	.60	.13	−.42	−1.34	−1.83
−1.0	1.59	1.31	1.12	.61	.16	−.38	−1.34	−1.87

the plotted relation of K to recurrence interval, T, for the Pearson Type III distribution for 5 values of C_s. Matalas (1963) describes the mathematical fitting process without use of a table of frequency factors. The computer program entitled "Revised Flood Statistics" is available for fitting a Pearson Type III curve to data.

An example of fitting a Pearson Type III curve by use of computed parameters and a table of frequency factors is given with the example of graphical fitting under the section of that name.

Type I extreme-value distribution (Gumbel)

This is a 2-parameter distribution having a constant skew of 1.139. The parameters are

$$u=\overline{X}-\bar{y}_N/\alpha$$

and

$$1/\alpha=S/\sigma_N,$$

where u is the mode, $1/\alpha$ is a scale parameter, $\overline{X}$ and S are the sample mean and standard deviation respectively, and $\bar{y}_N$ and σ_N are functions of N, the number of items in the sample. Values of $\bar{y}_N$ and σ_N for N from 8 to 1,000 are tabulated by Gumbel (1958, p. 228). Part of Gumbel's table is given in table 2.

The mean and standard deviation of the sample are computed, $\bar{y}_N$ and σ_N are read from the table, u and $1/\alpha$ are computed from the above formulas, and the straight line

$$X=u+y/\alpha \text{ is determined.}$$

This straight line is plotted on Powell-Gumbel probability paper using the "reduced variate y" scale. The Geological Survey Form 9-179a, flood data plot, does not have a reduced variate scale (but the recurrence interval is related to the reduced variate). On 9-179a plot the mean, $\overline{X}$, at the 2.33-year recurrence interval and use the approximate relation $y=ln\ T$ to locate another point on the straight line.

Following is a sample computation for annual floods on Columbia River near The Dalles, Oreg., for 1858-1946. See U.S. Geological Survey Water-Supply Paper 1080, page 337, for data.

Mean flood is 606,200 cfs$=\overline{X}$

Standard deviation, $S=\sqrt{(\sum X^2-N\overline{X}^2)/N-1}$
$=175,200.$

From table 2 for $N=89$,

$$\bar{y}_N=.558$$
$$\sigma_N=1.20,$$

then

$$1/\alpha=S/\sigma_N=175,200/1.20=146,000,$$

and

$$u=\overline{X}-\bar{y}_N/\alpha$$

$$=606,200-(0.558)(146,000)=524,700.$$

The equation of the straight line is

$$X=u+y/\alpha=524,700+146,000y.$$

The relation $y=ln\ T$ may be used to define plotting points for large recurrence intervals.

$$y=ln\ T=2.303 \log T.$$

For $T=50$ years, $y=3.91$,

and

$$X=524,700+146,000(3.91)=1,096,000 \text{ cfs.}$$

Table 2.—Means and standard deviations of reduced extremes

[Extracted from a more complete table by Gumbel]

N	$\bar{y}_N$	σ_N
10	0.4952	0.9497
15	.5128	1.021
20	.5236	1.063
25	.5309	1.091
30	.5362	1.112
35	.5403	1.128
40	.5436	1.141
45	.5463	1.152
50	.5485	1.161
60	.5521	1.175
70	.5548	1.185
80	.5569	1.194
90	.5586	1.201
100	.5600	1.206
200	.5672	1.236
500	.5724	1.259
1,000	.5745	1.269

The line is defined on the graph of figure 5 by the points

$$\overline{X}=606,200 \text{ cfs at 2.33 years}$$

and

$$1,096,000 \text{ cfs at 50 years.}$$

The plotted points for the period 1858-1948 are shown to indicate the fit of the computed line.

Type III extreme-value distribution

Examples of fitting this type of distribution to low-flow data are given by Gumbel (1954b).

Graphical Fitting

Graphical fitting requires no assumption as to the type or characteristics of the distribution. In the graphical method, each individual in the sample is assigned a probability or recurrence interval. Then magnitudes of the individuals are plotted against probabilities or recurrence intervals, and a line is drawn to properly interpret the points.

Assignment of probabilities is by means of a plotting-position formula. Various formulas may be used, each based on a different assumption as to the characteristics of the sample. Langbein (1960) relates the better-known plotting-position formulas to their underlying assumptions. Benson (1962b) compares the results of using various plotting positions on the economics of engineering planning. The Geological Survey uses the formula

$$T=1/p=(n+1)/m,$$

where T is recurrence interval in years, p is probability of an exceedence in any one year, n is the number of items in the sample, and m is the order number of the individual in the sample array (Dalrymple, 1960). Upper case symbols, P, N, and M are often used alternatively. The sample data may be arrayed—arranged in order of magnitude—beginning with the largest as No. 1, or with the smallest as No. 1, according to whether the frequency curve is to describe the probability of exceedence or of nonexceedence. A distribution curve can be cumulated from either end, and in the graphical method this effect is accomplished by selecting the direction in which the data are arrayed.

The next step is plotting magnitude against recurrence interval (or probability) on a graph. If arithmetic coordinates are used, an S-curve

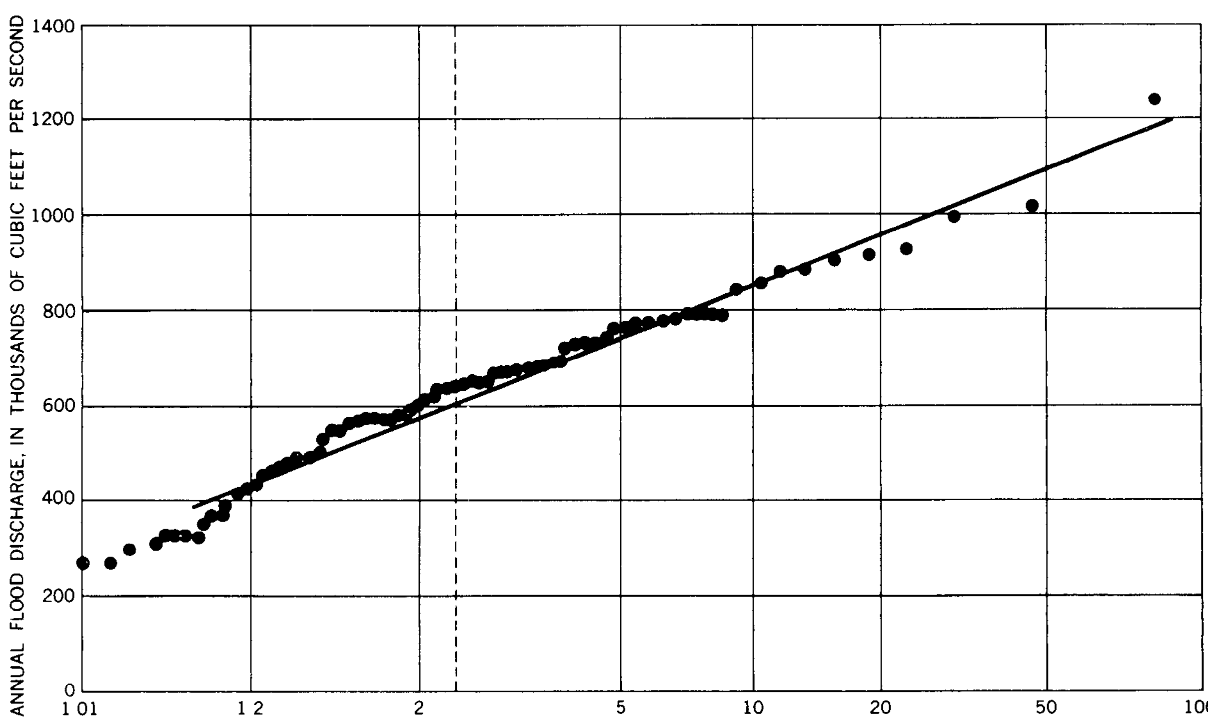

Figure 5.—Gumbel frequency curve of annual floods on Columbia River near The Dalles, Oreg., showing agreement with the plotted points.

usually results. It is difficult to define such a curve by the few observations; it is customary, therefore, to use a graph sheet having the abscissa graduated in such a way that a particular theoretical frequency curve will plot as a straight line. Such graph sheets are available for the normal, lognormal, and Gumbel Type I distributions. It is possible to prepare such a scale for any two-parameter distribution.

Although sets of data of the same type may not appear to lie on straight lines on a particular plotting paper, the lines of good fit usually are only slightly curved in one direction. Such lines may be more confidently defined from the plotted points than sharply curved lines. An additional advantage of the probability graph appears when a straight line is a reasonable interpretation of the plotted points; then the straight line is a frequency curve of the theoretical type on which the plotting paper is based. A discussion of normal-probability paper is given by Dixon and Massey (1957, p. 55–57). It should be clearly understood that a frequency curve is not necessarily normal just because the points are plotted on normal-probability paper (or has a Gumbel distribution because the points are plotted on Gumbel probability paper); only when the frequency curve is a straight line is this true.

The mean of a normal distribution corresponds to the 0.5 probability or to the 2-year recurrence interval. But a curved line on normal-probability paper represents a skewed distribution whose mean is not at 2-year recurrence interval. The effect of skew on the relation of mean to recurrence interval is easily demonstrated by use of the Gumbel Type I distribution which has a fixed positive skew. As used for flood analyses, the mean occurs at 2.33-year recurrence interval. But if the same Gumbel distribution is used to represent the frequency of floods less than, the positions of the mean and median are reversed, and the mean plots at about 1.59 years. This effect is shown by figures 3 and 4. The discharge corresponding to the 2.33-year recurrence interval as obtained from a curved line on Gumbel probability paper is not the mean. It can, however, be used as a characteristic discharge as could the 2-year value or any other near the central part of the distribution.

Example of graphical fitting

The annual discharges for the years 1915–50 inclusive in table 3, column 2, can be used to define a frequency curve. The curve can be cumulated from the high end or from the low end, depending on whether the data are arrayed from the high end or from the low end. Both arrays are given in table 3.

Table 3.—Computation of plotting position

Water year	Q	Order number, m; highest as No. 1	Plotting position $(n+1)/m$	Order number m; lowest as No. 1	Plotting position $(n+1)/m_1$
1915	264	34	1.09	3	11.2
16	374	11	3.37	26	1.43
17	332	19	1.95	18	2.06
18	346	16	2.31	21	1.76
19	359	13	2.85	24	1.54
1920	333	18	2.06	19	1.95
21	483	3	11.2	34	1.09
22	417	5	7.40	32	1.16
23	346	17	2.18	20	1.85
24	320	21	1.76	16	2.31
1925	271	31	1.19	6	6.16
26	214	36	1.03	1	37.0
27	530	2	18.5	35	1.06
28	304	25	1.48	12	3.09
29	271	32	1.16	5	7.40
1930	271	33	1.12	4	9.25
31	304	26	1.43	11	3.37
32	400	9	4.11	28	1.32
33	327	20	1.85	17	2.18
34	415	6	6.16	31	1,19
1935	402	8	4.62	29	1.28
36	362	12	3.09	25	1.48
37	320	22	1.68	15	2.47
38	272	30	1.23	7	5.30
39	244	35	1.06	2	18.5
1940	279	28	1.32	9	4.11
41	303	27	1.37	10	3.70
42	310	24	1.54	13	2.85
43	275	29	1.28	8	4.62
44	317	23	1.61	14	2.65
1945	350	15	2.47	22	1.68
46	387	10	3.70	27	1.37
47	359	14	2.65	23	1.61
48	449	4	9.25	33	1.12
49	406	7	5.30	30	1.23
1950	570	1	37.0	36	1.03

Arraying 20 or 25 items, that is, arranging them in order of magnitude, and assigning order numbers, can be done readily by observation. For a larger number of items, various schemes may be used. One method is to write each item and its year of occurrence on a card, then arrange the cards in order of magnitude, number the cards, and transfer the order numbers to the table of items. Another method utilizes transparent plastic strips, one for each period of record used. Each strip has a calibrated length equal to the abscissa scale on Geological Survey Form 9–179a or 9–179b. On the strip are marked the plotting positions for the particular n, identified by order number. Thus, it is unnecessary to compute plotting positions by this procedure, and plotting can be done rapidly. Most attractive of all is a machine program developed by the U.S. Bureau of Public Roads; this program produces the probability graph with the points plotted on it.

The plotting positions given in table 3 are estimated recurrence intervals; probabilities would be their reciprocals; n is the number of items ($1950-1915+1=36$), and m is the order number. Plotting positions may be computed by slide rule or read from a table prepared for the purpose.

Discharge is plotted against recurrence interval on Geological Survey Form 9–179a or 9–179b. The former has an arithmetic ordinate scale, and the latter a logarithmic ordinate scale. The abscissa scales are alike and are based on the Gumbel Type I distribution. Figures 6 and 7 show plots of the data from table 3. Flood data are usually plotted on Form 9–179a, and minimum-flow data on Form 9–179b. The points are averaged by eye in drawing the lines, with the extreme points given less weight than the others because their recurrence intervals are not so well defined.

Interpretation of the curves would be as follows:

On figure 6 the discharge defined by the curve at 10-year recurrence interval would be exceeded as an annual maximum at intervals averaging 10 years in length. Conversely, figure 7 indicates that the discharge at intervals averaging 10 years in length will be less than the curve value corresponding to the 10-year recurrence interval. Further interpretation of a frequency curve is given on page 12.

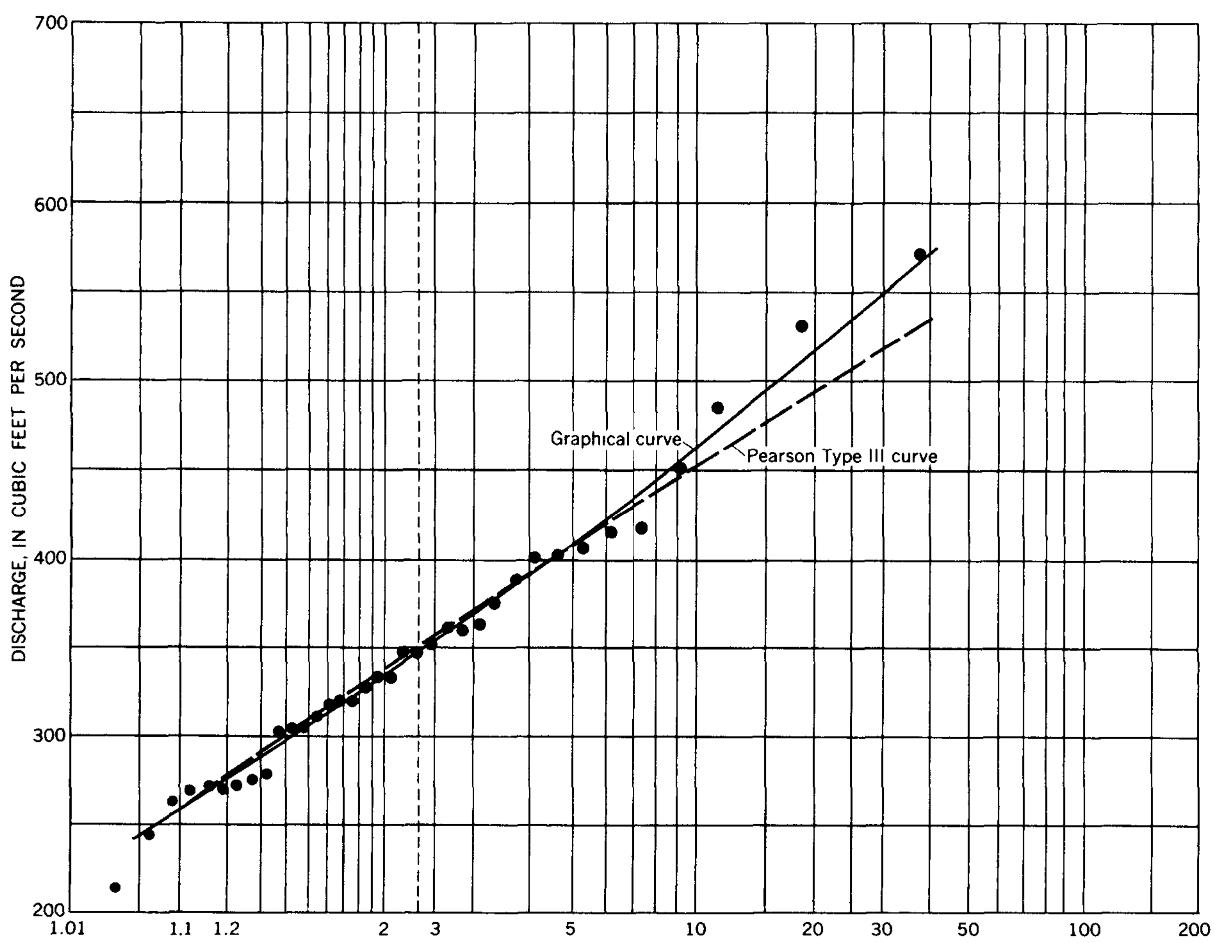

Figure 6.—Frequency curve based on data from table 3, assuming that data are annual maximums.

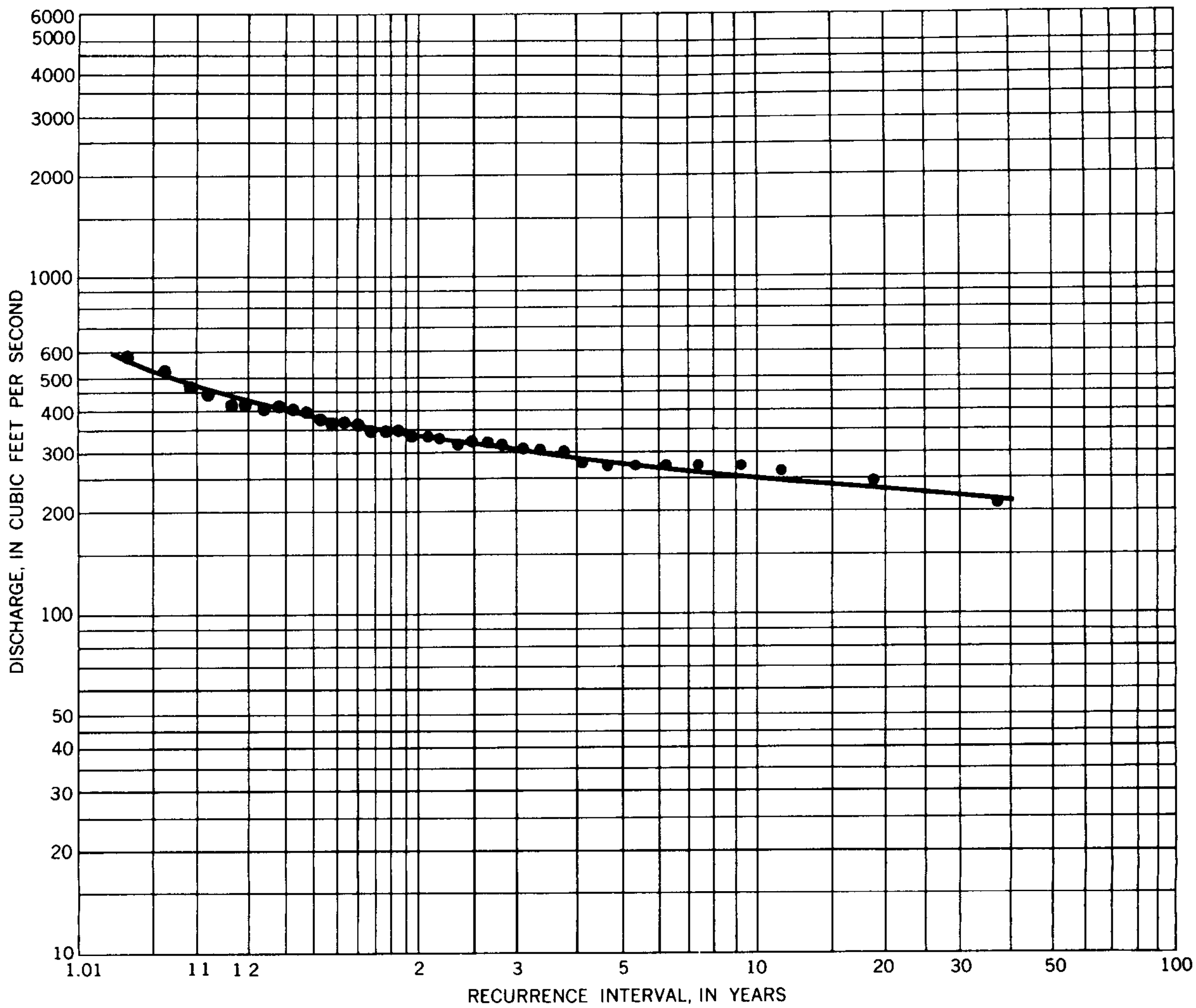

Figure 7.—Frequency curve based on data from table 3, assuming that data are annual minimums.

For comparison with the graphical curve of gure 6, the theoretical Pearson Type III curve defined by the data of table 3 is computed as follows:

$$\text{Mean} = \overline{Q} = \sum Q/N = 12{,}486/36 = 347$$

$$\text{Variance} = S^2 = \frac{\sum Q^2 - (\sum Q)^2/N}{N-1}$$

$$= \frac{4{,}542{,}500 - 155{,}900{,}196/36}{35} = 6055$$

$$\text{Standard deviation} = S = 77.8$$

$$\text{Coefficient of skew} = C_s = \frac{N^2 \sum Q^3 - 3N \sum Q \sum Q^2 + 2(\sum Q)^3}{N(N-1)(N-2)S^3}$$

$$= \frac{(36)^2(1{,}738{,}665{,}756) - 3(36)(12{,}486)(4{,}542{,}500) + 2(12{,}486)^3}{(36)(35)(34)(6055)(77.8)}$$

$C_s = 1.04.$

A simpler definition of C_s is the ratio of the third moment about the mean to the 3/2 power of the second moment about the mean; that is

$$C_s = \frac{\sum (Q-\overline{Q})^3/N}{S^3},$$

which, for this example, gives $C_s=0.94$ if S is not adjusted by the factor $N/N-1$. The difference between results of the two formulas for C_s will increase as N decreases.

Plotting positions are computed by

$$Q=\overline{Q}+KS,$$

where $\overline{Q}$ and S are as computed above, and K is obtained from table 1 for various recurrence intervals at a C_s of 1.0. For example, for the 20-year recurrence interval,

$$Q=347+1.87(77.8)=492.$$

It can be seen from table 1 that the frequency factor is little affected by changes of a tenth or so in the coefficient of skew; therefore, a simple method of computing C_s should be adequate. The theoretical Pearson Type III curve defined by the data is plotted on figure 6.

Computation of the coefficient of skew requires cubing the individual values, which is time consuming on a desk computer. Theoretical fitting, therefore, is most easily done on a digital computer.

Use of historical data

The graphical method of defining a frequency curve is readily adaptable to inclusion of certain historical data. Essentially, an estimate of the recurrence interval of each historical event is made on the basis of available information, and the event plotted at this recurrence interval. Occasionally, the historical information may indicate the need to modify the recurrence interval of a flood within the period of record. Dalrymple (1960, p. 16–18) describes the procedure.

Comparison of Mathematical and Graphical Fitting

Mathematical fitting has the following advantages:

1. For the same theoretical distribution, every analyst using a given set of data would get the same answer.
2. Fitting can be done by electronic computer.
3. The result can be completely described by a few parameters.

Mathematical fitting has the following disadvantages:

1. Selection of the theoretical distribution is arbitrary.
2. No one theoretical distribution will fit all data of one type such as flood peaks.
3. Characteristics of a set of data tend to be obscured if they are not plotted; however, plotting can be done separately.
4. No objective method is available for incorporating historical information in the computation.

Graphical fitting provides the following advantages:

1. The procedure is simple and can be done quickly.
2. No assumption as to the particular form of the distribution need be made.
3. Relation of the curve to the points is readily seen.
4. Historical data may be included.

Graphical fitting has the following disadvantages:

1. Even though the same plotting position formula is used, different analysts will draw somewhat different frequency curves.
2. The result cannot be described by two or three parameters.

The above comparisons of mathematically and graphically fitted frequency curves are based on statistical considerations. If there were one underlying distribution for a particular streamflow characteristic, such as annual flood peaks, then fitting that distribution mathematically to all sets of annual flood-peak data would be desirable. Under that condition it is assumed that the type of population distribution is known, and the sample is used to estimate the parameters. Among several basins of equal size the differences in computed parameters would be due to sampling errors only. But basins are not only not the same size, they differ also in their flood-generating characteristics and, consequently, would have different population frequency distributions. Thus, the variability among a group of annual flood-peak frequency curves is due to sampling error and differences in basin characteristics; any advantage of mathematical fitting is reduced because physical characteristics may produce a frequency curve that is unlike any theoretical one. The large influence of basin characteristics on the shape of certain low-flow frequency curves is described by Riggs (1965). Likewise, a flood-frequency curve for a stream draining a basin composed of a humid mountainous part and a semiarid lowland part would be a composite having an irregular shape not closely represented by any theoretical frequency distribution.

Describing Frequency Curves

It is sometimes desired to characterize frequency curves by means of numerical indexes for comparing several curves or for use in hydrologic analyses. The mean or median is a commonly used index of central tendency. Variability is usually described by the coefficient of variation, $C_v=S/\overline{X}$, which is the standard deviation divided by the mean, and thus is dimensionless.

The mathematical fitting process provides values of the mean and standard deviation. If a three-parameter distribution is fitted, the coefficient of skew is also computed. These three parameters completely describe the distribution.

The mean and standard deviation of graphically fitted frequency curves are readily obtained from the graph if the frequency curve is a straight line on normal or lognormal probability paper.

The usual graphically fitted frequency curve is not a straight line on the plotting paper used; consequently, the mean is not at a known probability or recurrence interval, the standard deviation cannot be accurately determined, and the curvature indicates the existence of skewness. For such curves it is customary to use the median flow as the characteristic of central tendency. The Lane-Lei (Lane and Lei, 1950) variability index may be used to describe the variability. The Lane-Lei index is an approximation of the standard deviation and is computed as follows from a plot on lognormal probability paper:

1. Values of discharge at 10 percent intervals from 5 to 95 percent are read from the curve (a probability scale rather than a recurrence interval scale is used).
2. The logarithms of these values are found.
3. The standard deviation of the logarithms $\sqrt{\sum (\log X-\overline{\log X})^2/N-1}$ is the variability index.

A coefficient of variation can also be defined as the variability index divided by the median.

It is not customary to estimate the skewness of a graphical frequency curve because skewness is of little use in characterizing such a curve.

More specific comparisons of frequency curves can be made by considering magnitudes at particular probability levels or recurrence intervals. The mean is such a one, of course. Others commonly used are the annual minimum flow at 2-year recurrence interval and flood peaks at many recurrence intervals (Benson, 1962a).

Interpretation of Frequency Curves

A frequency curve based on random homogeneous data is an estimate of the cumulative probability distribution of the population from which the sample was drawn. The following interpretations of the frequency curve require the assumption that the curve is a good representation of the population distribution.

Referring to the graphical curve of figure 6, the recurrence interval of 500 cfs is 16 years. This means that the annual maximum will exceed 500 cfs at intervals averaging 16 years in length, or that the probability of the annual maximum exceeding 500 cfs in any one year is 1/16.

From figure 7 the recurrence interval of 250 cfs is 13 years. Thus, the annual minimum discharge will be less than 250 cfs at intervals averaging 13 years in length, and the probability that the minimum discharge in any one year will be less than 250 cfs is 1/13.

Many interpretations of frequency curves in hydrology have been stated in terms of the probability "of equaling or exceeding" a selected value. Most variables in hydrology, notably streamflow, are continuous—but reported as discrete—and the theoretical probability of occurrence of any particular value in a continuous distribution is zero. Therefore it seems desirable to use only "of exceeding."

The interpretations of frequency curves given above will not answer questions such as the probability of an event of 10-year recurrence interval being exceeded in a 10-year period. Intuitively, one might expect that probability to be 0.5, but it is not. The correct probability can be computed as follows. Since the probability of exceeding the 10-year event in 1 year is 0.1, the probability of not exceeding it in 1 year is 0.9. Then, the probability of not exceeding it in 10 years is, by the multiplicative law of probabilities

$$(0.9)^{10}=0.35,$$

and the probability of one or more events exceeding the 10-year event in 10 years is 0.65.

A more complete interpretation of a frequency curve is given by Riggs (1961) who also proposed in that reference, that the frequency curve be used as a basis for a family of curves giving the probability of events exceeding certain magnitudes in given periods of years (design periods). Figure 8 shows a flood-frequency curve and the design-probability curves computed from it.

It should be noted that the above discussion does not apply to frequency curves based on the Beard (1943) method of computing plotting positions. For such curves the n-year event has a probability of 0.5 of not being exceeded in n years. See Langbein (1960) for a further discussion of Beard's plotting position.

Although frequency curves are used as though they were accurate representations of the population distribution, we know that they may not be. Benson (1960) sampled from a known distribution and showed a wide range in shape and position of frequency curves defined by different samples of the same size. Another way of assessing the reliability of a frequency curve is by computing the confidence limits. Chow (1964, p. 8–31) describes a method. These computations indicate that the frequency curve is most reliable in the vicinity of the mean.

In comparing frequency curves for two different streams, the analyst should keep in mind

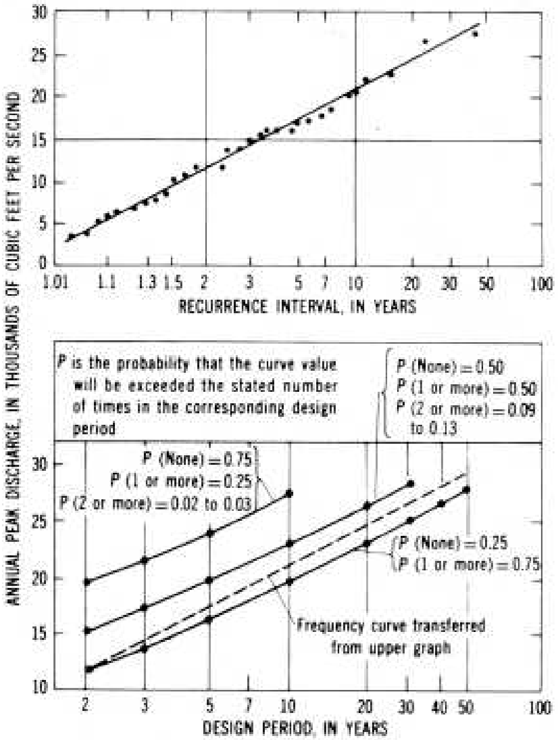

Figure 8.—Design-probability curves (lower graph) and the frequency curve on which they are based (upper graph).

that the curves may differ because of chance or because of the effects of different basin characteristics, or both. The population distribution of a flow characteristic of one stream may be considerably different from the population distribution of that characteristic for another stream. Riggs (1965) discusses some basin characteristics which influence the shape of the frequency curve of annual minimum flows.

Special Cumulative Frequency Curves

Frequency curves are often useful for defining the distribution of events even though the events are not entirely independent of each other (that is, they are serially correlated), in which case the probability interpretation must be somewhat modified and cannot be precisely stated.

An example of a frequency curve based on serially correlated data is that of the annual minimum flows of South Fork Obion River,

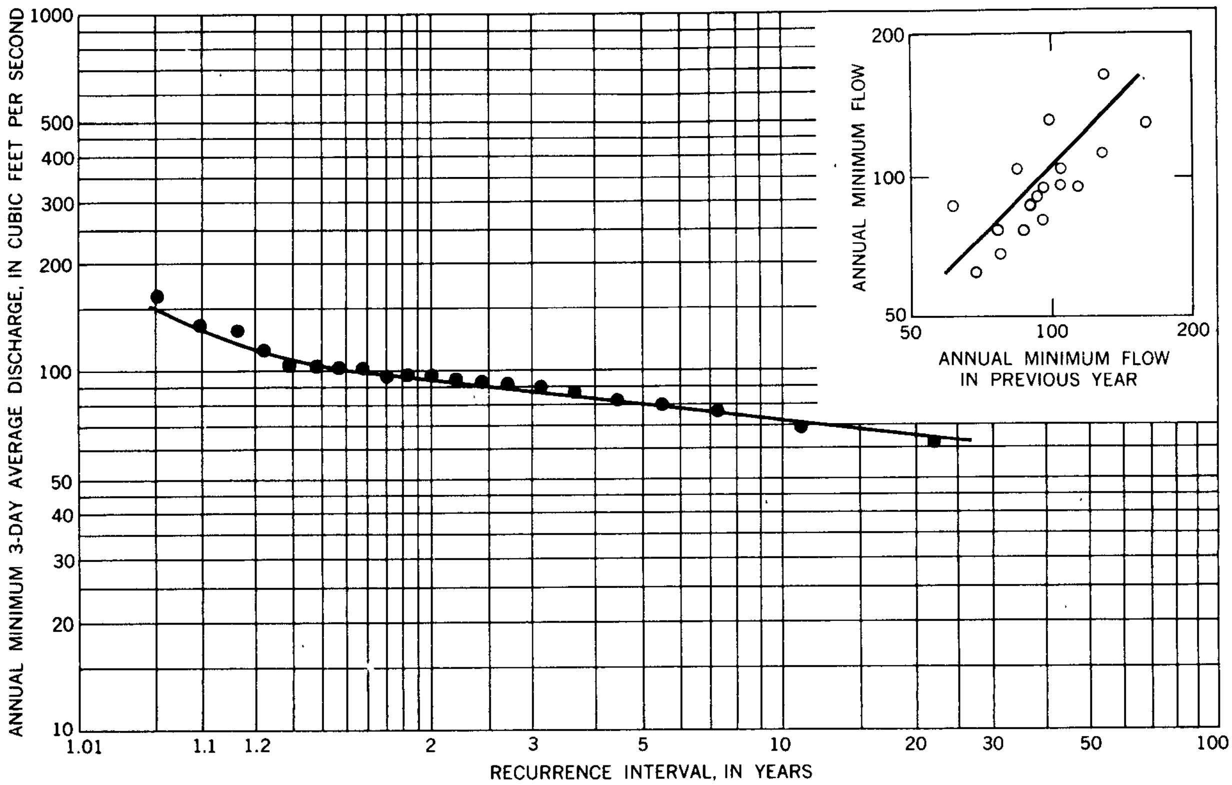

Figure 9.—Frequency curve of annual minimum flows and plot showing serial correlation, South Fork Obion River near Greenfield, Tenn.

Tenn., in figure 9. Also in figure 9 is shown the first-order serial correlation between the annual minimums. The existence of this serial correlation should warn the analyst that interpretations of the frequency curve in the usual way will be subject to more than the usual uncertainty.

The frequency distribution of daily mean discharges of a stream, plotted on a log normal graph, is called a duration curve (Searcy 1959). Daily mean discharges are not only serially correlated, they are nonhomogeneous because of the yearly cycle in streamflow which produces different means and ranges of discharge at different times of the year. Thus, the duration curve cannot be interpreted as a probability curve, but it is useful as a description of the distribution of daily means that has occurred and may be expected to recur over a period of several years. The duration curve also has other uses because its shape and position are defined by basin characteristics.

Although the annual flood series is commonly used for frequency analysis, the partial-duration series is also used. This series is made up of all floods above an arbitrary base regardless of their time sequence. Such a series is not a true statistical series and cannot be treated rigorously. Use of the term "recurrence interval" with the partial-duration series introduces difficulties if the series includes more floods than years. The partial-duration flood-frequency curve is useful in studying the frequency of inundation. Dalrymple (1960) describes current practice. Riggs and Thomas (1965) discuss the interpretation and use of the method.

Frequency curves of mean discharge for periods longer than 1 year have been used in storage analyses (Stall and Neill, 1961; Stall, 1964). Such curves must be interpreted carefully; the recurrence interval in years of an event that extends over a period of more than 1 year is hard to visualize except for the No. 1 item in the array.

The foregoing special frequency curves provide some information on probability of exceedence of an individual event, although this probability may not be estimated directly or precisely. Another type of frequency curve is prepared merely to describe the distribution of events by size; probability does not enter into the problem because the distribution of future events is not of interest. A well-known curve of this type is one showing the distribution of particle sizes of a soil or sediment sample.

References Cited

Beard, L. R., 1943, Statistical analysis in hydrology: Am. Soc. Civil Engineers Trans., v. 108, p. 1110–1121.

——— 1962, Statistical methods in hydrology: U.S. Army Engineer District, Corps Engineers, Sacramento, Calif.

——— 1964, Estimating long-term storage requirements and firm yield of rivers: Internat. Assoc. Sci. Hydrology, General Assembly of Berkeley, Pub. 63, p. 151–166.

Benson, M. A., 1960, Characteristics of frequency curves based on a theoretical 1,000-year record, *in* Dalrymple, Tate, Flood frequency analyses: U.S. Geol. Survey Water-Supply Paper 1543–A, 80 p.

——— 1962a, Factors influencing the occurrence of floods in a humid region of diverse terrain: U.S. Geol. Survey Water-Supply Paper 1580–B, 64 p.

——— 1962b, Plotting positions and economics of engineering planning: Am. Soc. Civil Engineers Proc., v. 88 HY 6 p. 57–71, Paper 3317.

Chow, V. T., 1954, The log-probability law and its engineering applications: Am. Soc. Civil Engineers Proc., v. 80, separate 536, p. 9.

——— 1964, Frequency analysis, *in* Chow, V. T., Handbook of applied hydrology: New York, McGraw-Hill Book Co., p. 8–17,

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

Dixon, W. J., and Massey, F. J. 1957, Introduction to statistical analysis, 2d ed.: New York, McGraw-Hill Book Co.

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

——— 1954a Statistical theory of extreme values and some practical applications: U.S. Dept. Commerce, Natl. Bur. Standards Applied Mathematics Series 33, p. 12.

Gumbel, E. J., 1954b, Statistical theory of droughts: Am. Soc. Civil Engineers Proc. v., 80, separate 439, 19 p.

——— 1958, Statistics of extremes: New York, Columbia Univ. Press, 375 p.

Hazen, Allen, 1930, Flood flows, a study of frequencies and magnitudes: New York, John Wiley & Sons, Inc.

Lane, E. W., and Lei, Kai, 1950, Stream flow variability: Am. Soc. Civil Engineers Trans. v. 115, p. 1084–1098.

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

Matalas, N. C., 1963, Probability distribution of low flows: U.S. Geol. Survey Prof. Paper 434–A, 27 p.

Mood, A. M., 1950, Introduction to the theory of statistics: New York, McGraw-Hill Book Co., Inc.

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

Riggs. H. C., 1953, A method of forecasting low flow of streams: Am. Geophys. Union Trans., v. 34, no. 3, p. 427–434.

——— 1961, Frequency of natural events: Am. Soc. Civil Engineers Proc., v. 87, no. HY1, p. 15–26.

——— 1965, Estimating probability distributions of drought flows: Water and Sewage Works, v. 112. no. 5, p. 153–157.

——— 1967, Some statistical tools in hydrology: U.S. Geol. Survey Tech. Water-Resources Inv., book 4, chap. A1 (in press).

Riggs, H. C., and Thomas, D. M., 1965, Discussion of "Mathematical model for flood risk evaluation," by Richard M. Shane and Walter R. Lynn: Am. Soc. Civil Engineers Proc., v. 91, no. HY5, p. 190–192.

Searcy, J. K., 1959, Flow-duration curves: U.S. Geol. Survey Water Supply Paper 1542–A, 33 p.

Stall, J. B., 1964, Low flows of Illinois streams for impounding reservoir design: Illinois State Water Survey Bull. 51, 395 p.

Stall, J. B., and Neill, J. C., 1961, A partial duration series for low flow analyses: Jour. Geophys. Research, v. 66, p. 4219–4225.

☆U S GOVERNMENT PRINTING OFFICE· 1989-242-384/93973